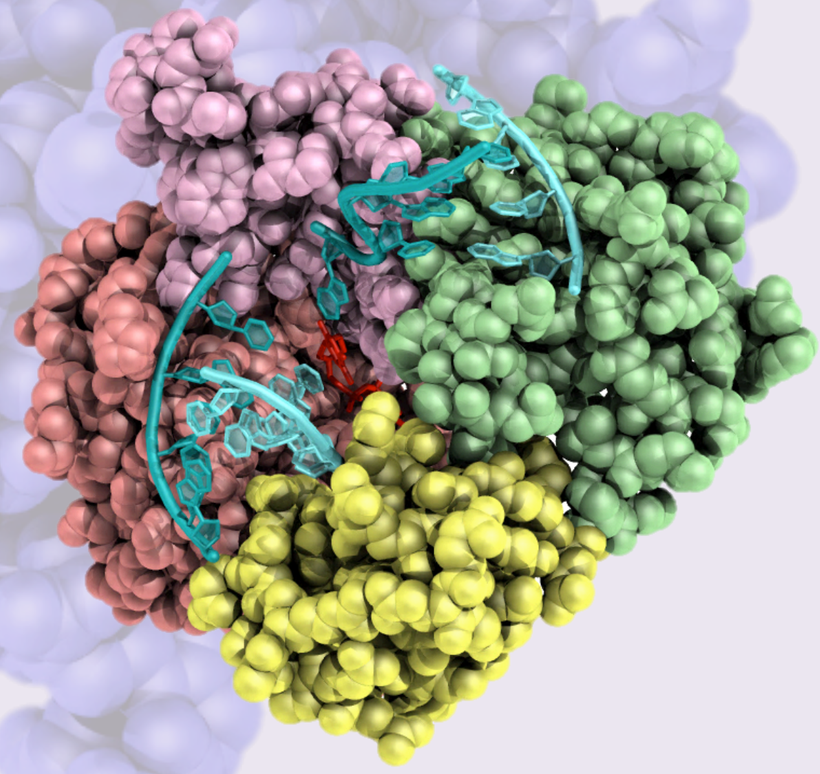
Maria José Martín Pereira

TESIS DOCTORAL

UAM
UNIVERSIDAD AUTONOMA
DE MADRID
Facultad de Ciencias
Departamento de Biología Molecular

# EXCLUSIVE POLYMERASES REPAIRING DNA BREAKS

## THE SAME MAGICS FROM BACTERIA TO MAN



Madrid
2011

**Maria José Martín Pereira**
Madrid, 2011

UNIVERSIDAD AUTONOMA DE MADRID

DEPARTAMENTO DE BIOLOGÍA MOLECULAR
FACULTAD DE CIENCIAS

# EXCLUSIVE POLYMERASES REPAIRING DNA BREAKS

# THE SAME MAGICS FROM BACTERIA TO MAN

MARIA JOSE MARTIN PEREIRA
LICENCIADA EN BIOQUÍMICA

MADRID 2011

DEPARTMENT OF MOLECULAR BIOLOGY
FACULTY OF SCIENCES


# EXCLUSIVE POLYMERASES REPAIRING DNA BREAKS

# THE SAME MAGICS FROM BACTERIA TO MAN

MARIA JOSE MARTIN PEREIRA
Biochemistry Licentiate

SUPERVISOR: PROF. LUIS BLANCO

MADRID 2011

Memoria presentada por María José Martín Pereira, Licenciada en Bioquímica, para optar al Grado de Doctora en Ciencias por la Universidad Autónoma de Madrid.

Thesis submitted by María José Martín Pereira, Biochemistry Licentiate, in order to qualify for the degree of Doctor of Philosophy in Sciences by the Universidad Autónoma de Madrid.

La presente Tesis Doctoral ha sido realizada bajo la supervisión del Dr. Luis Blanco Dávila, Profesor de Investigación del Consejo Superior de Investigaciones Científicas, adscrito al Centro de Biología Molecular Severo Ochoa.

This PhD Thesis has been performed under the supervision of Dr. Luis Blanco, Research Professor of the Spanish Research Council (CSIC), at the “Severo Ochoa” Molecular Biology Centre.

Fdo.: Luis Blanco Dávila

*A mis padres.*
*A Jose.*

*Si es de alguien, es de ellos.*

*There is no first world and third world.*
*There is only one world,*
*for all of us to live and delight in.*

*(No hay primer mundo y tercer mundo.*
*Sólo hay un mundo,*
*para que todos vivamos y nos deleitemos en él.)*

Gerald Durrell

*I love fools' experiments.*
*I am always making them.*

*(Me encantan los experimentos locos.*
*Siempre los hago.)*

Charles Darwin

# *Agradecimientos*

Éste es uno de los momentos más especiales que he vivido: un paso adelante enorme después de mucho tiempo de trabajo, experimentos, esfuerzo, viajes, penas y alegrías. Por eso quiero compartirlo y agradecérselo a todas las personas que me han empujado hasta este punto, unas veces a sabiendas y otras casi sin darse cuenta.

En primer lugar, al Dr. Luis Blanco Dávila, por aceptarme en su grupo para realizar la Tesis Doctoral. El querer pertenecer a su laboratorio ha sido una de las mejores decisiones que he tomado, ya que he aprendido muchísimo tanto de la mitad práctica de la ciencia como de la mitad teórica, que yo considero la más importante. Gracias, Luis, por ser un maestro paciente e incansable, un jefe exigente pero comprensivo y cercano, por apoyarme en todo momento y en todas mis iniciativas, y, sobre todo,

Quiero agradecer a la Dra. Margarita Salas el haber sido mi tutora durante estos casi cinco años, y a todo su laboratorio la ayuda que me prestaron desde el principio de mi tesis: a Laura, que me enseñó a hacer ***footprints*** cuando casi ni sabía lo que era marcar un oligo, a Josemari, que me enseñó a hacer kilos de fosfocelulosa nueva, entre otras muchas cosas, a Miguel de Vega, por ser como eres y hacérnoslo pasar tan bien en Suiza y aquí, y a todos por crear buen ambiente en el ala larga de la planta cuarta.

A mis compañeros del laboratorio, que fue 122 en el antiguo CV y ahora es 403 en nuestro nuevo CBM: Raquel, Ángel, Gloria, Arancha, Paula, Francisco, Mariví, Verónica, Susana, María, Sara, Ana Aza, Ana Gómez, Guille, Jose y Sandra, quiero agradecerles lo bien que lo he pasado estos años. Por todos los buenos momentos, que estaban llenos de risas, algún que otro cotilleo, canciones de Disney (¡estamos de viernes!), conversaciones interesantes en la comida y mucha alegría. Vero, gracias por esos desayunos que nos ayudaban a mediar la mañana con mejor cara, y por incluirme en tu vida de esa manera tan especial. Que sigamos viendo crecer esta amistad y, de momento, ¡a tu niño! Mariví merece una mención especial por atreverse a hablarnos en Oxford y, después de eso, venir al laboratorio, por haberme ayudado y enseñado lo que son el tesón y la constancia incluso aunque yo no estuviera, por hacerme reír casi hasta las lágrimas a cada momento y por ser como es, ¡cordobesa loquilla!

Quiero agradecerle muy especialmente a la Dra. Maria Julia Marinissen el haberme permitido aprender tanto de ella durante un periodo que fue, para mí, demasiado corto. Maria Julia, ya sabes lo que me hubiese gustado que nuestra aventura continuara, pero esos pocos meses te fueron suficientes para dejar tu huella en todas las personas que te conocieron de cerca, yo entre ellas. Gracias por enseñarme a trabajar como se debe en un laboratorio, siendo yo aún una estudiante de la universidad, por ser mi primera mentora científica, siempre disponible para resolver dudas y mirar ensayos de luciferasa y westerns, y por ser además una amiga a corta y larga distancia. Creo sinceramente que yo sería muy distinta de no haberte conocido. Por supuesto no me olvido de mi compañera de laboratorio en el IIB durante esa etapa, Ana, que ayudó mucho a que aquel tiempo fuese más divertido, y con la que sigo charlando por los pasillos del CBM. También quiero recordar aquí a Tamara, argentina adorable enamorada de Madrid, que nos visitó durante unos meses y que supo hacerse un huequito que se sigue manteniendo, aunque no hayamos aún logrado vernos en Suiza. Y a Carla, que conseguía alumbrar ese pasillo de la Facultad de Medicina y que sigue pisando con fuerza en esta profesión nuestra, tan esquiva a veces.

Maria Julia, generosamente me ofreciste la oportunidad de viajar al otro lado del océano y aprender a desenvolverme en otro ambiente y con otros compañeros, y así conocí a mucha gente a la que recuerdo con cariño: Dani, Rebeca, Chepe, Adrian, Vyomesh, Todd, Ann, Lynn, Bob, Julie, y todos los compañeros del laboratorio del Dr. Silvio Gutkind. A él también le deseo agradecer el haber permitido que fuese de visita durante dos meses del verano de 2006, y que accediese a que su post-doc de mi barrio madrileño, Dani, perdiese un poco el tiempo ayudándome a encontrarme en aquel laboratorio del NIH con tantas ***rooms***.

Con este primer viaje me entró el gusanillo y no he parado, entre estancias y congresos, con la ayuda, unas veces, de esas becas que parece que no existen y luego se encuentran y consiguen, y otras de Luis, que siempre ha estado a favor de que me fuese a otros laboratorios para aprender, sí, pero no sólo ciencia, sino también experiencia. Gracias a esto quiero creer que he madurado un poco, he recibido una

educación más completa durante mi tesis y he conocido a muchos amigos que siguen ayudándome cuando lo necesito, y haciéndome reír y asombrarme de vez en cuando con sus idas y venidas.

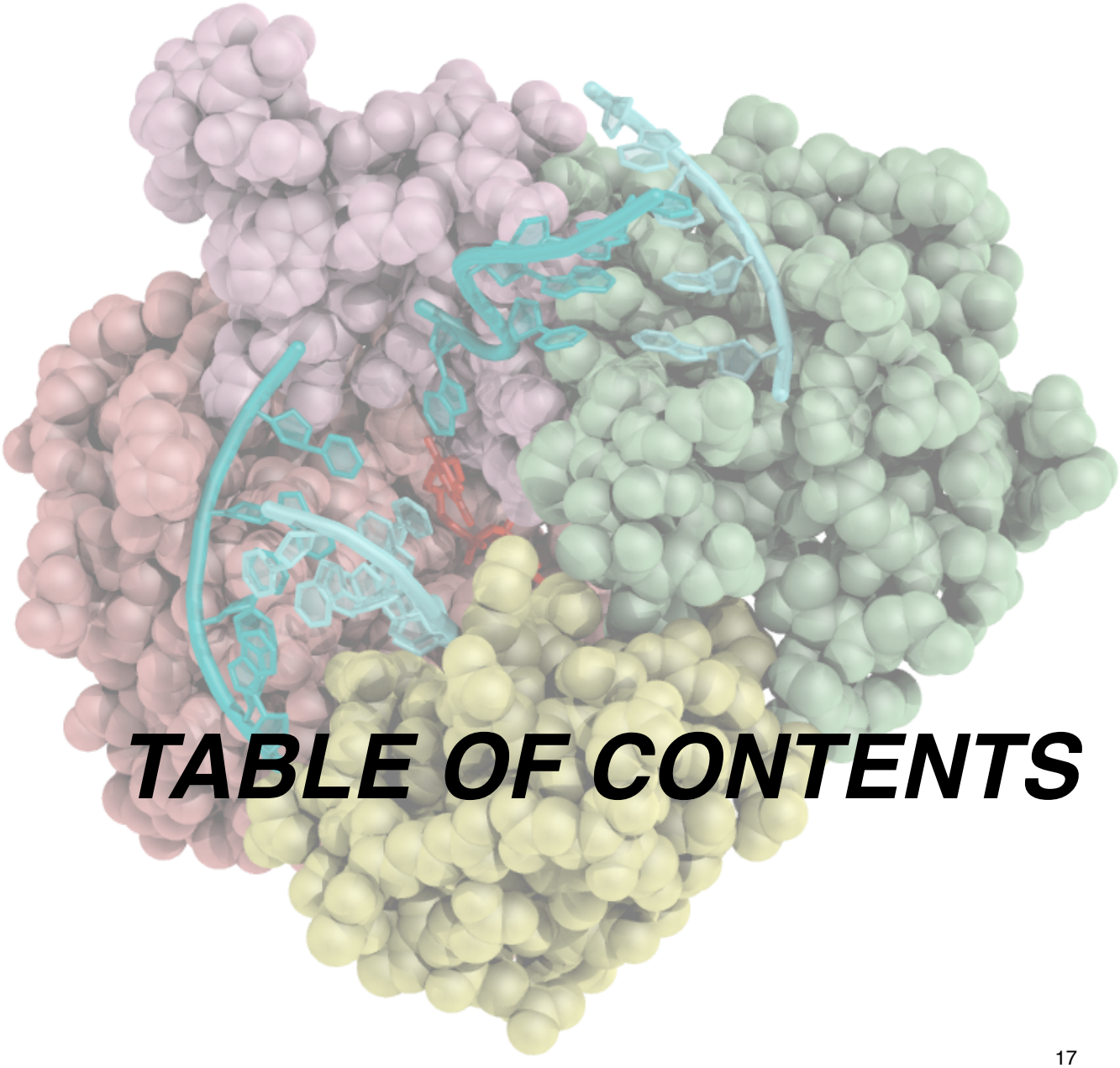
Le agradezco al Dr. Aidan Doherty que me acogiese en su laboratorio del soleado Brighton (incluso en otoño), y quiero enviarles mi cariño a todos los chicos de su laboratorio, que me trataron como una más durante aquellos 3+1 meses (¡dichoso gato!): Nigel, Ed, Andy, Tamsyn, Julie, Sean, Pierre y Helen. Conocí mucha más gente durante este tiempo, y quiero recordar especialmente a la "Spanish people": Felipe y María y su Dieguito, Queti y su familia, y Elena, con quien pasé muchos fines de semana de turismo y algunas horas angustiosas en el tren.

Mi segunda visita a Estados Unidos, esta vez en Stony Brook, fue perfecta, en primer lugar por haber escogido el laboratorio del Dr. Miguel García-Díaz, un "Blanquito" de aúpa, al que agradezco mucho el haberme aceptado no sólo en el trabajo sino también en su casa. Él y su familia me hicieron pasar algunos de los mejores ratos allí. Gracias a los tres por esas tardes de animadas charlas, de juegos infantiles y de música. Además, en el laboratorio tuve la suerte de estar rodeada de gente estupenda: Edison, Elenas H. y Y., James, y lo que parecían hordas de estudiantes que querían aprender de Miguel, como yo, el arte de hacer "joyitas" de proteína. Miguel, eres un profesor estupendo con la increíble capacidad de hacerte comprender, hasta explicando las cosas más complicadas en comandos de linux. Mis chicos Edison y Óscar, ya sabéis todo lo que os quiero y lo que os recuerdo, vosotros más que nadie conseguisteis que ese verano de 2010 fuese inolvidable para una Marilyn morena y española, entre días en la playa, salidas de compras y musicales de Broadway.

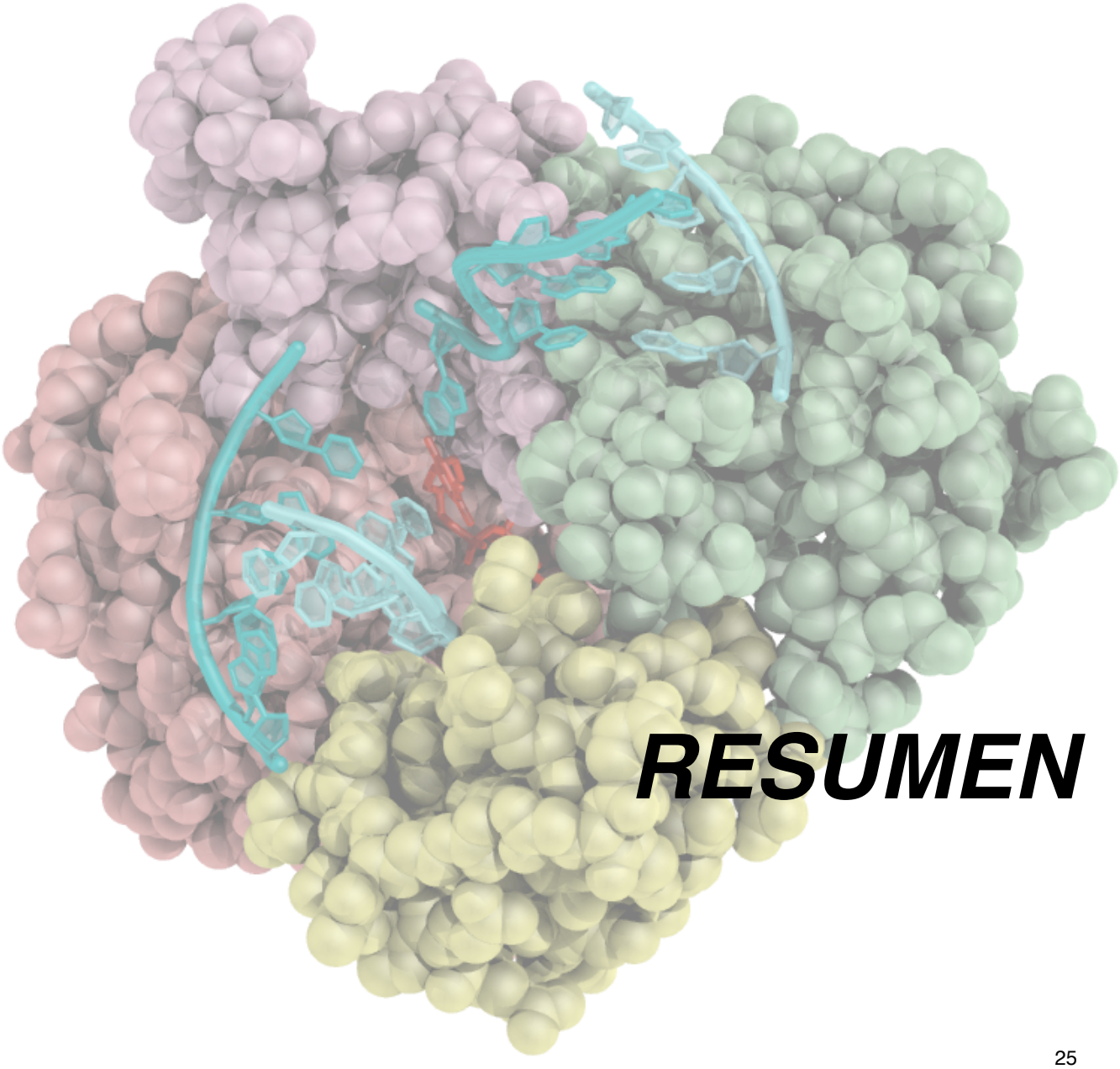
A mis niñas, Erica, Marta, María y Adriana, gracias por escucharme cuando lo he necesitado, por hacerme reír con vuestras locuritas, por las vacaciones de verano e invierno, en ciudades o en la playa (¡y en la India!), por enseñarme de arquitectos y edificios (y de muchas otras cosas, también de la vida), por querer oírme hablar de evolución y de proteínas, por vuestra alegría siempre, por permitirme formar parte de vuestras vidas, ¡por quererme!. Y a los hombres, Carlos Bonet y Serrano, Julio y Alexis (¡ay, chico!), por llenar de diversión y de muchas cosas interesantes cada momento que hemos pasado juntos. Hemos visto como todos íbamos cambiando y subiendo escaloncitos, quiero agradeceros el haberme ayudado a subir éste. Gracias a todos por ser mis amigos, que lo sigamos siendo y nuestros niños lo sean también. ¡Os quiero!

Mis padres, durante toda mi vida, han sido ejemplos de comportamiento a los que querría parecerme más. Siempre me han dado libertad y me han apoyado en todo lo que he querido hacer, sin perder de vista el inculcarme la mejor educación y valores. Ellos son los responsables de todo lo bueno que pueda haber en mí. Con respecto a esta Tesis, ellos han sido el soporte indispensable sin el cual nada hubiese sucedido, sumando las "cosas de mamá" (los tuppers de cada día, la ropita limpia) y las "locuras de papá" (el humor tan "granaíno", los viajes de repente), pero, sobre todo, el amor que hace que estemos los tres tan unidos.

Y a mi Jose, que me soporta hasta cuando no me soporto ni yo, que me ayuda siempre con toda su inteligencia y fortaleza, que me guía por el camino que vamos proyectando entre los dos, que se deja llevar casi siempre por mis locuras pero consigue frenarme cuando debe, que es feliz a mi lado y me hace feliz al suyo... por todo esto y todo lo que queda por llegar, ¡gracias!



# TABLE OF CONTENTS



# RESUMEN

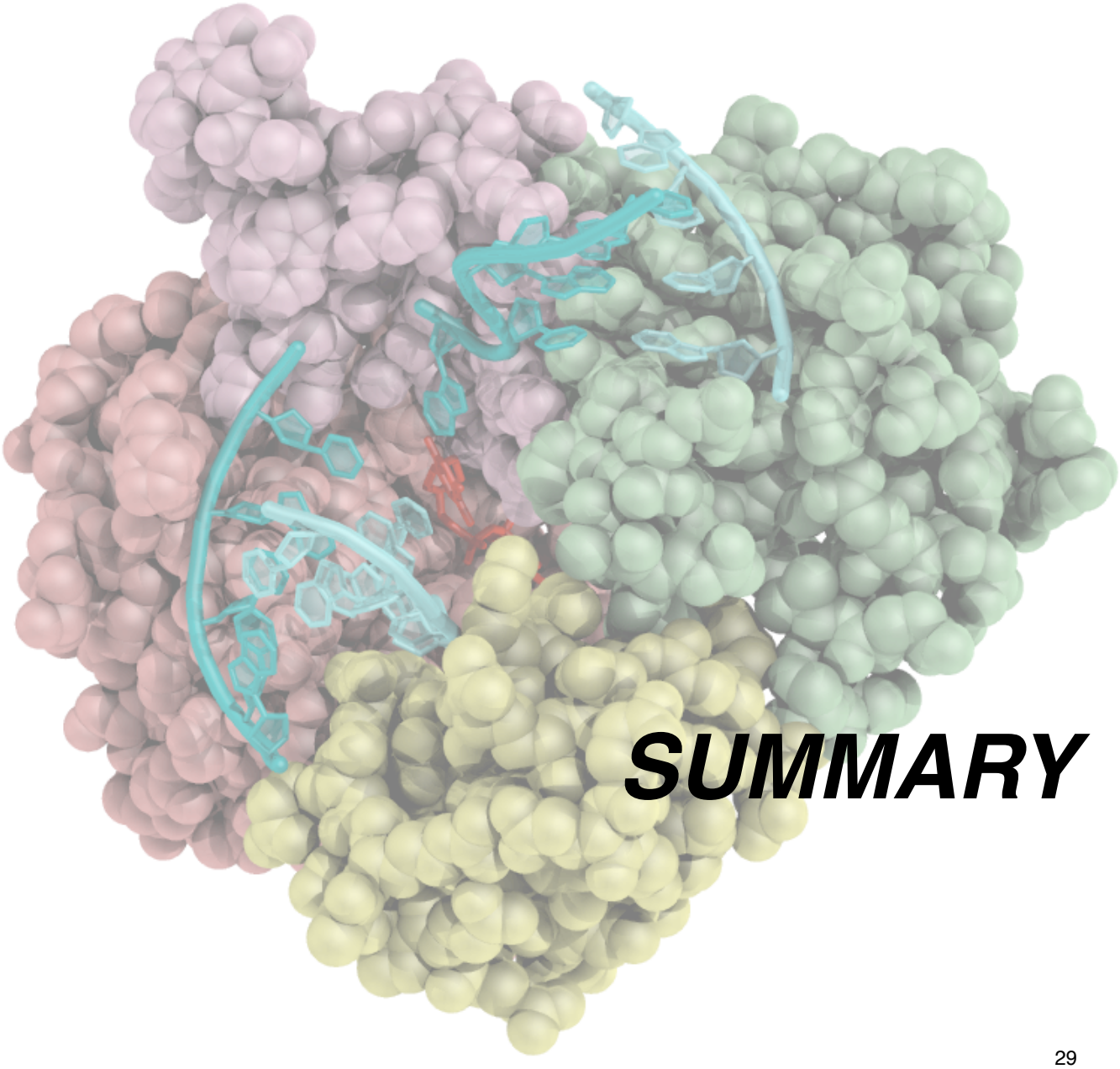
La DNA polimerasa *μ* (Pol*μ*), miembro de la familia X de enzimas de reparación de DNA, está equipada con actividades de polimerización y transferasa terminal en el mismo centro activo de la polimerasa. Tal flexibilidad hace de esta enzima un candidato idóneo para la creación de variabilidad genómica, pero también para contribuir a la reparación de roturas de doble cadena por unión de extremos no homólogos (NHEJ). Además, Pol*μ* es capaz de unir dos extremos de DNA en ausencia de factores accesorios de NHEJ, con los que interactúa a través de su dominio BRCT, lo que sugiere su participación en la vía alternativa de NHEJ. En este trabajo se muestra que el dominio BRCT también está directamente implicado en la unión más estable de Pol*μ* a los sustratos de DNA, y por lo tanto en una mayor eficiencia de la reacción.

Nuestros estudios también arrojan luz sobre el mecanismo por el cual una transferasa terminal limitada en Pol*μ* podría minimizar la generación de variabilidad en NHEJ, especialmente al reparar extremos incompatibles. Así, el análisis de la información 3D disponible y nuestros resultados sugieren que el paso limitante de la transferasa terminal implica el movimiento del extremo de la cadena *primer* (3'-protuberante) desde un complejo E:DNA no productivo a un complejo E:DNA:dNTP productivo, para lograr la adición de nucleótidos al azar, con la participación del residuo $Arg^{387}$ como un regulador/limitador específico. El cambio en Pol*μ* del modo “transferasa terminal” al modo “polimerización dictada por molde” en las reacciones de NHEJ es provocado por el reconocimiento del grupo 5’-P en el segundo extremo de DNA, cuya protuberancia 3’ podría proporcionar una base molde que facilitase la translocación del *primer*, aliviando así el paso limitante de la reacción.

Ademas de compartir la habilidad de insertar nucleótidos independientemente de molde, otra similitud entre TdT y Pol*μ* es su promiscuidad en el uso de ribonucleótidos (NTPs), además de dNTPs, cuya importancia fisiológica es actualmente desconocida. Por ello, hemos evaluado la relevancia del uso de NTPs como sustratos para NHEJ de extremos incompatibles, situación que podría tener un impacto asociado en mutagénesis, identificando las condiciones que conducen a un aumento en la eficiencia sin un coste en términos de fidelidad mediante el uso de NTPs y varias combinaciones de metal activador. Hemos demostrado por primera vez que la fidelidad de la enzima se mantiene cuando los iones $Mn^{2+}$ se utilizan en la concentración fisiológica adecuada. Proponemos que Pol*μ* podría tomar ventaja de un sitio activo versátil, válido para la selección alternativa de diversos iones metálicos y nucleótidos como sustratos. Esta versatilidad le permite la selección "ad hoc" de la combinación nucleótido/metal más adecuada para cada evento de NHEJ, ampliando así el espectro de soluciones disponibles en función de la orientación de la base molde durante la conexión.

Mediante experimentos de *footprint* mostramos que la unión de Pol*μ* a estructuras de DNA tipo *gap* es más eficiente que la de Polλ, otro miembro de la familia X que también participa en NHEJ, pero requiere un mínimo de complementariedad de los dos extremos. En este sentido, hemos demostrado la importancia para el NHEJ de dos características estructurales, presentes en Pol*μ* y también en TdT, como son el Loop 1 y el motivo YACQR (“broche”), posiblemente implicado en el mantenimiento de una conformación cerrada del núcleo de la polimerasa. También se han caracterizado varios residuos implicados específicamente en la unión de los dos extremos de DNA a través de contactos con las cadenas *primer* y *template*.

En muchos procariotas, una primasa arqueo-eucariota (AEP) constituye el dominio polimerasa (PolDom) de una proteína multifuncional, LigD, implicada en la reparación de roturas del DNA mediante NHEJ. En esta tesis doctoral se presenta la estructura del PolDom de *Mycobacterium tuberculosis* en una conformación sin precedentes: el complejo pre-ternario. Este complejo, catalíticamente competente, consta de un monómero PolDom, un extremo del DNA 3’-protuberante, dos iones metálicos ($Mn^{2+}$) y el dNTP complementario a la base molde, pero, significativamente, carece de una cadena *primer*. El análisis por mutagénesis dirigida ha demostrado que el aminoácido invariante $Arg^{220}$, localizado en el loop2 y que es específico para la AEPs bacterianas que participan en NHEJ, juega un importante papel en la catálisis mediante la regulación de la unión de un ion metálico en el sitio activo. Tal complejo pre-ternario y pre-activado de la polimerasa se produce en solución, y puede utilizar hasta un dinucleótido como *primer* mínimo, lo que podría ser ventajoso para las reacciones NHEJ de extremos 3'-protuberantes muy cortos o no complementarios, y puede facilitar también actividades como la transferasa terminal. En esta tesis doctoral también se estudia una segunda estructura, que muestra por primera vez una sinapsis totalmente complementaria de dos extremos de DNA que incluye un dímero de Mt-PolDom. Nuestros resultados de mutagénesis dirigida muestran que el loop2 no sólo participa en la catálisis, sino también en el posicionamiento correcto del *primer*. Además, la correcta selección de la base molde durante el NHEJ resulta definida por dos fenilalaninas, que proporcionan a PolDom la alternativa de copiar correctamente el molde o generar cambios en el marco de lectura (*frameshifts*), consiguiéndose así la versatilidad necesaria. Las similitudes entre el sistema de NHEJ humano y el bacteriano, con respecto a las polimerasas participantes, se discuten en profundidad a la luz de los resultados presentados.



# SUMMARY

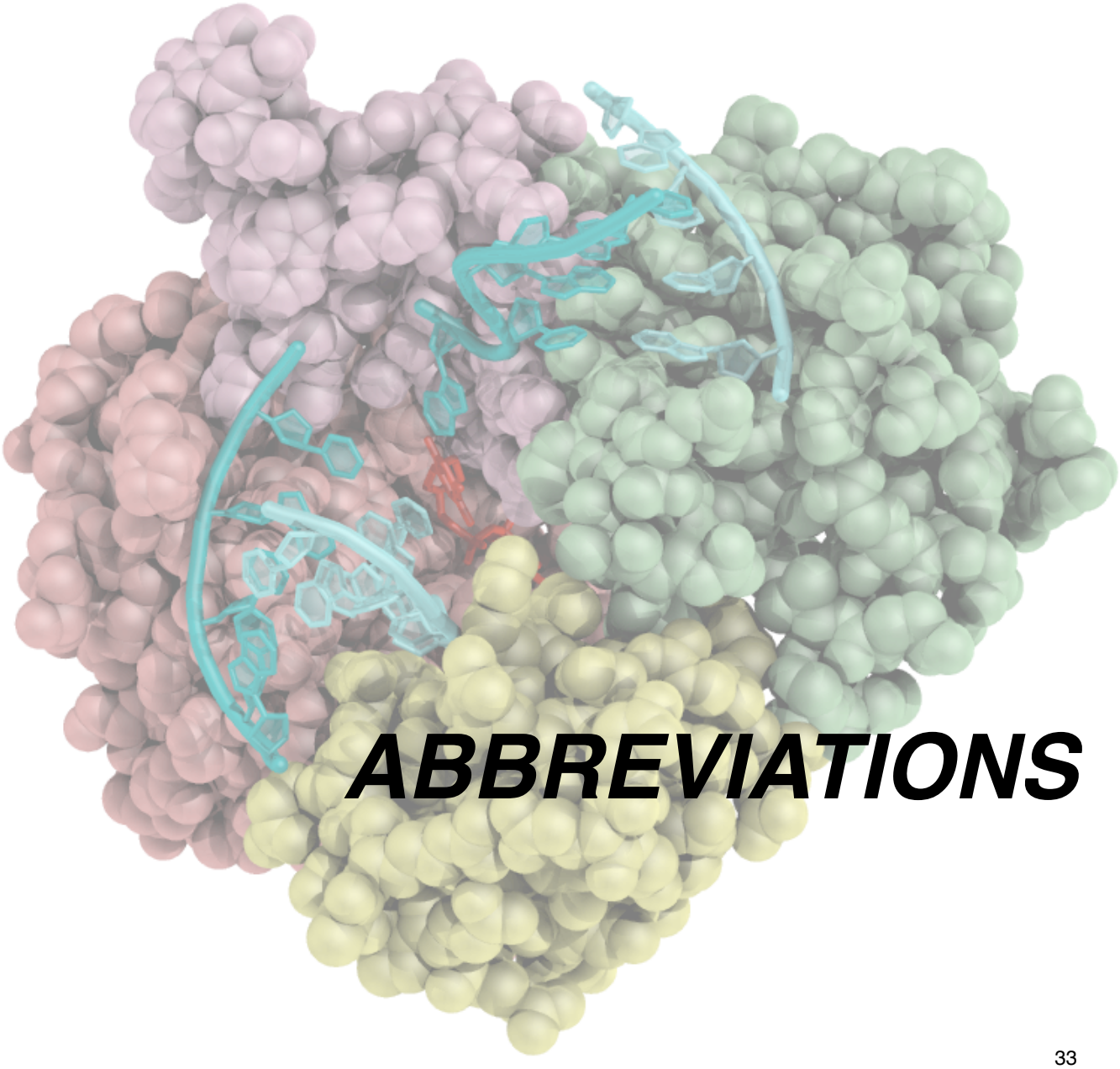
DNA polymerase *μ* (Pol*μ*) is a family X member involved in DNA repair, equipped with both template directed and terminal transferase activities in the same polymerase core. Such a flexibility makes this enzyme a candidate to create genomic variability, but also to contribute to end-joining of non-complementary ends (NHEJ). Its wide tissue distribution is compatible with compelling data supporting a general role in DNA repair of DSBs. Also, Pol*μ* is able to join two DNA ends in the absence of any NHEJ accessory factors, to which Pol*μ* interacts *via* its BRCT domain, suggesting its participation in the alternative end-joining pathway. In this work we show that the BRCT domain is also involved in promoting a tighter binding of Pol*μ* to the DNA substrates, and thus an increased efficiency.

Our studies also shed light on the mechanism by which a rate limited terminal transferase activity in Pol*μ* could avoid generation of unnecessary variability during NHEJ, specially when dealing with incompatible ends: analysis of the 3D information available and our results suggest that the rate-limiting step of the terminal transferase most likely involves the movement of the single-stranded (3'-protruding) primer-terminus from a nonproductive E:DNA complex to a productive E:DNA:dNTP complex, to achieve untemplated addition of nucleotides, involving $Arg^{387}$ as a specific regulator residue. The Pol*μ* switch from terminal transferase to templated insertions in NHEJ reactions is triggered by recognition of a 5'-P at a second DNA end, whose 3'-protrusion could provide a templating base to facilitate such a special "pre-catalytic translocation step".

Apart from sharing the ability to incorporate untemplated nucleotides, another similarity between TdT and Pol*μ* is their promiscuity in using ribonucleotides in addition to dNTPs, whose physiological significance is presently unknown. We have evaluated the relevance of using NTPs as substrates for NHEJ of incompatible ends, a situation that could have an impact in associated mutagenesis, and we have identified the conditions that would lead to a gain in efficiency without a cost in terms of fidelity by using ribonucleotides and several activating metal ion combinations. Activating $Mn^{2+}$ ions, which provide a tighter binding of incoming nucleotides and are known to be optimal activators of Pol*μ*'s terminal transferase, do benefit NHEJ by improving the efficiency of nucleotide insertion. We have proved for the first time that the accuracy of the enzyme is maintained when $Mn^{2+}$ ions are used at the appropriate physiological concentration. We propose that Pol*μ* could take advantage of a versatile active site, valid for selecting alternative activating metal ions and nucleotides as substrates. This versatility would allow "ad hoc" selection of the most appropriate nucleotide/metal combination for individual NHEJ events, thus widening the spectrum of available solutions to position a templating base in proper register for connection.

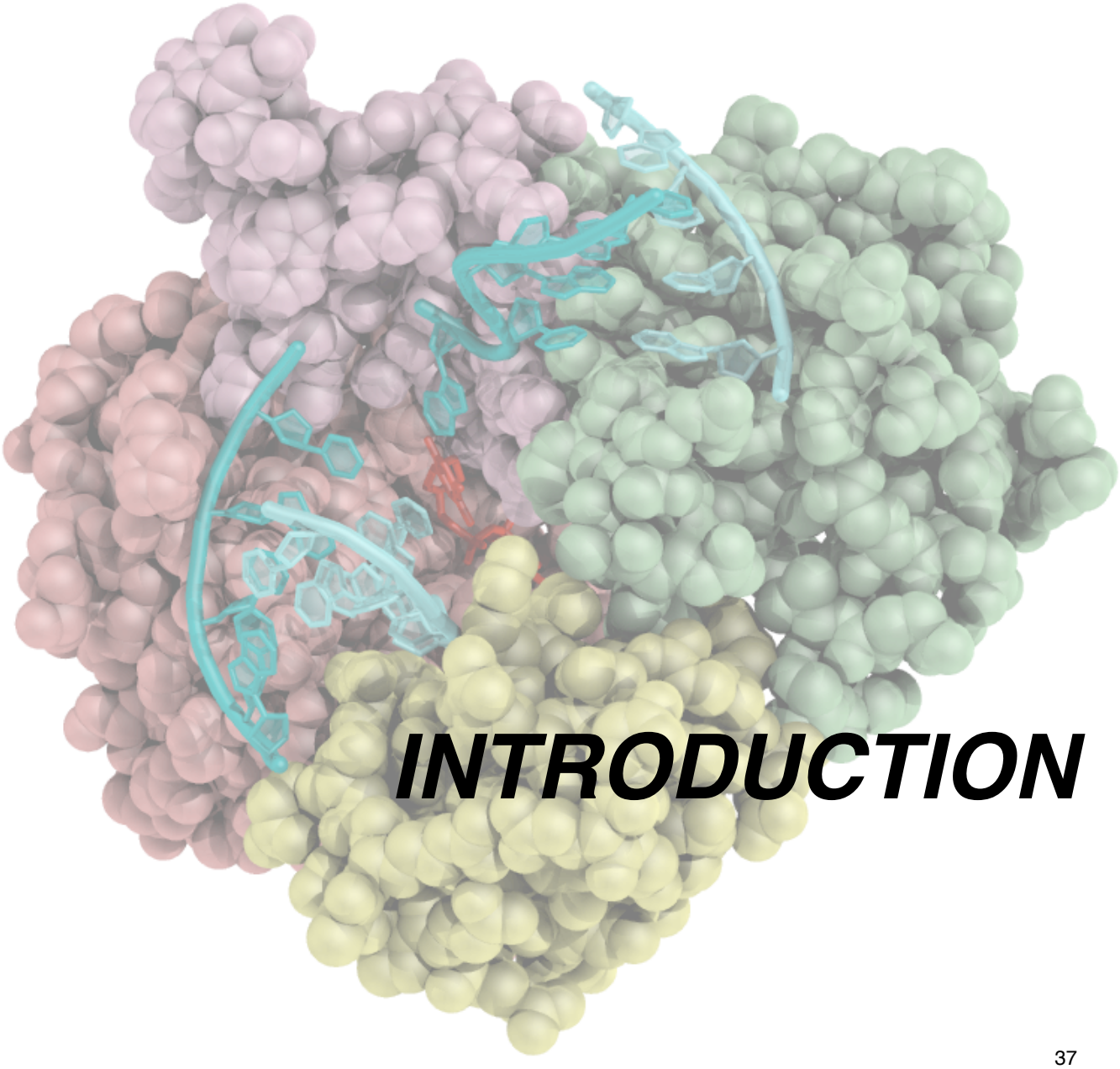
By DNA footprinting experiments we show that Pol*μ* binds gapped DNA structures more efficiently than Polλ, a related PolX member which also participates in NHEJ, but requires a minimal complementarity of the two ends to be intrinsically handled by the polymerase. In this regard, we have demonstrated the importance for NHEJ of several structural features, present in Pol*μ* and also in TdT, such as Loop 1 and the specific YACQR motif ("brooch"), potentially implicated in the maintenance of a closed conformation of the Polß-like core of the polymerase. Other structural features described in this PhD Thesis include several residues specifically involved in bridging the two DNA ends *via* contacts with the primer and the template strands.

In many prokaryotes, an Archaeo-Eukaryotic Primase (AEP) constitutes the polymerase domain (PolDom) of a multifunctional protein, LigD, involved in the repair of DNA DSBs by NHEJ. In this PhD thesis, a catalytically active conformation of *Mycobacterium tuberculosis* PolDom, which represents an unprecedented pre-ternary complex, is presented. This catalytically competent complex consists of a PolDom monomer, a DNA end with a 3''overhang, two metal ions ($Mn^{2+}$) and an incoming nucleotide complementary to the templating base but, significantly, it lacks a primer strand. This structure represents a unique example of a polymerase-DNA complex in a pre-ternary intermediate state. Site-directed mutational analysis demonstrated that the invariant $Arg^{220}$, contained in the conserved loop2, that is specific for the bacterial AEPs involved in NHEJ, plays an important role in catalysis by regulating binding of a second metal ion at the active site. Such a pre-activated polymerase intermediate state, shown to occur in solution and able to use even a dinucleotide as the minimal primer, could be advantageous to facilitate NHEJ reactions involving critically short 3'-protruding or non-complementary DNA ends and may also facilitate unusual extension activities associated with these enzymes, such as terminal transferase reactions. A second structure, showing for the first time a fully complementary synapsis of two DNA ends achieved by a *Mt*-PolDom dimer, is also studied in this PhD Thesis. Our site-directed mutagenesis results show that loop2 is not only involved in catalysis but also in the correct accommodation of the incoming primer. Also, correct selection of the templating base during NHEJ is shown to rely on two phenylalanines, that supply PolDom with the ability either to correctly copy the template or to generate frameshifts, thus providing the required versatility. The similarities between the human and bacterial NHEJ system, with regard to the polymerases involved, are discussed in depth in light of the results presented.



# ABBREVIATIONS

| | |
|---|---|
| 8oxoG | 7,8-dihydro-8-oxoguanine |
| BER | Base excision repair |
| BRCT | BRCA1 C-terminal domain |
| DSB | Double strand break |
| IPTG | Isopropyl-ß-D-thiogalactopyranoside |
| KO | Knock-out |
| NHEJ | Non homologous end-joining |
| NLS | Nuclear localization signal |
| nt | Nucleotide |
| P | Phosphate |
| Polß | DNA polymerase beta |
| Polλ | DNA polymerase lambda |
| Pol*μ* | DNA polymerase mu |
| TdT | Terminal deoxynucleotidyl transferase |



# INTRODUCTION

The cells of all living organisms have their genetic information stored as DNA. This information should remain unchanged and be transmitted to daughter cells upon cell division, to allow maintenance and proper functioning of organisms. This is carried out at the molecular level by DNA polymerases, enzymes that perform DNA synthesis and are responsible for two fundamental processes in the cell: replication and repair of the DNA molecule. Replication is the process that is aimed at duplicating the genetic material, allowing its transmission from one generation to the next. Repair processes help maintaining the integrity of DNA, so that genetic information remains stable and unchanged, as the DNA is susceptible to damage by multiple agents, both endogenous and exogenous. In both processes there is always the possibility of introducing mutations, which, although they may cause cell death or lead to tumorigenesis, are also key to generate variability in certain programmed processes or to fuel gene evolution.

## 1. DNA Polymerases

The model of Watson and Crick explained the structure of DNA: an antiparallel double helix of two complementary strands. In this model, they also glimpsed the idea of a possible DNA copying mechanism (Watson and Crick, 1953). Later in 1958, Kornberg's group discovered the DNA polymerase I from *E. coli,* the first member of this class of enzymes (Lehman et al., 1958). Demonstrating its ability to catalyze DNA synthesis according to the instructions on the template strand was key to understanding how DNA is synthesized within the cell, and provided the basis for many subsequent studies that began to reveal further complexities of the DNA replication mechanism. So far, several enzymes of this type have been identified and characterized in diverse organisms. DNA polymerases are involved in not only DNA replication and repair, but also in other mechanisms of damage tolerance, recombination and mutagenesis.

Irrespective of their detailed domain structures, all polymerases whose structures are known share a common overall architectural feature: they have a shape that can be compared with that of a right hand consisting of “thumb,” “palm,” and “fingers” domains (Ollis et al., 1985; Fig. 1). The function of the palm domain appears to be catalysis of the phosphoryl transfer reaction whereas that of the fingers domain includes important interactions with the incoming nucleoside triphosphate as well as the template base to which it is paired. The thumb may play a role in positioning the duplex DNA and in processivity and translocation. The exception is ASVF PolX, which lacks the “fingers” domain (Oliveros et al., 1997; Showalter et al., 2001; Maciejewski et al., 2001).

Apart from these basic features, DNA polymerases have a broad diversification. Due to the complexity and diversity of the processes of replication and repair, polymerases are often highly specialized, with wide variability at the level of structure and biochemical functions. While the structure of its active site is usually preserved, the rest of the molecule may vary considerably and can have different additional domains bearing other enzymatic activities such as the exonuclease, lyase or primase among others, or non-catalytic domains to interact either with other protein factors or with DNA. Biochemically, DNA polymerases differ in the number of nucleotides added to a 3’-hydroxyl end in a single association event (distributive: a few nucleotides; processive, hundreds to thousands of nucleotides), in their efficiency to carry out the polymerization reaction, or in their preference for a specific type of substrate (large regions of ssDNA as

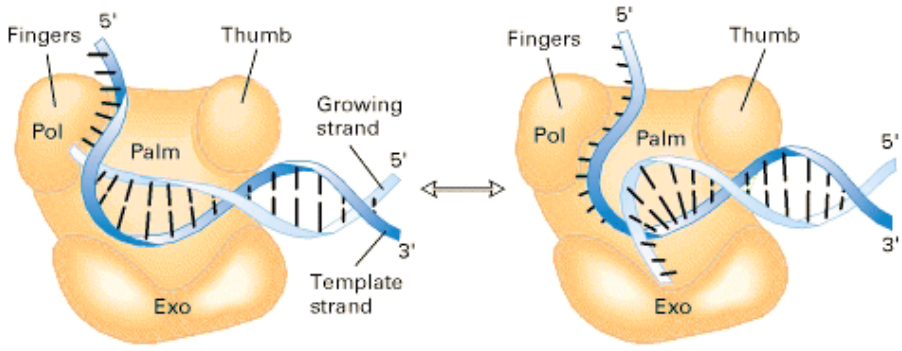



**Figure 1. Replicative polymerases.** Schematic representation of a replicative polymerase (orange) with exonuclease (exo) and polymerase (pol) domains, showing the similarities between the three sub-domains of the pol and a right hand: fingers, palm and thumb. The double stranded DNA substrate is shown in blue.

| Greek name | HUGO name | Family | Proposed function |
|---|---|---|---|
| **Polγ (gamma)** | POLG | A | Replication/Repair of mitochondrial DNA |
| **Polθ (theta)** | POLQ | A | TLS/SHM |
| **Polν (nu)** | POLN | A | TLS |
| | | | |
| **Polα (alpha)** | POLA | B | Nuclear replication/HR |
| **Polδ (delta)** | POLD1 | B | Nuclear replication/HR/MMR/NER |
| **Polε (epsilon)** | POLE | B | Nuclear replication/HR/MMR/NER |
| **Polζ (zeta)** | POLZ | B | TLS |
| | | | |
| **Polη (eta)** | POLH | Y | TLS |
| **Polι (iota)** | POLI | Y | TLS |
| **Polκ (kappa)** | POLK | Y | TLS |
| | REV1 | Y | TLS |
| | | | |
| **Polβ (beta)** | POLB | X | BER |
| **Polλ (lambda)** | POLL | X | Meiosis/ BER/NHEJ |
| **Polμ (mu)** | POLM | X | NHEJ/SHM/V(D)J |
| | TdT | X | V(D)J |

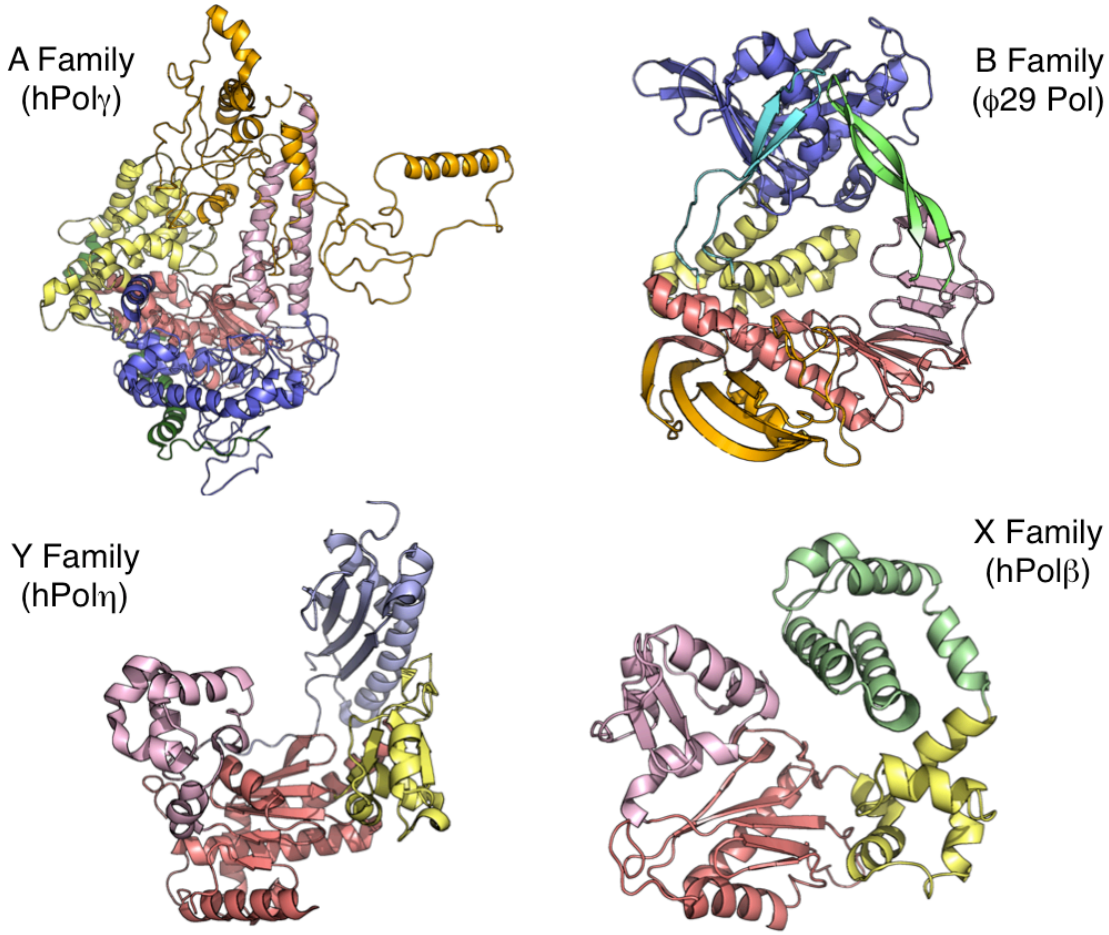



**Figure 2. Families of eukaryotic polymerases.**
Cartoon representation of the crystal structure of representative members of the four families of eukaryotic polymerases: Polγ from the A family, ϕ29 from B family, Polη from the Y family and Polß from the X family. General colouring shows fingers in yellow, palm in red, thumb in pink. Specific domains of each family are also coloured in orange, light or dark green and light or dark blue.

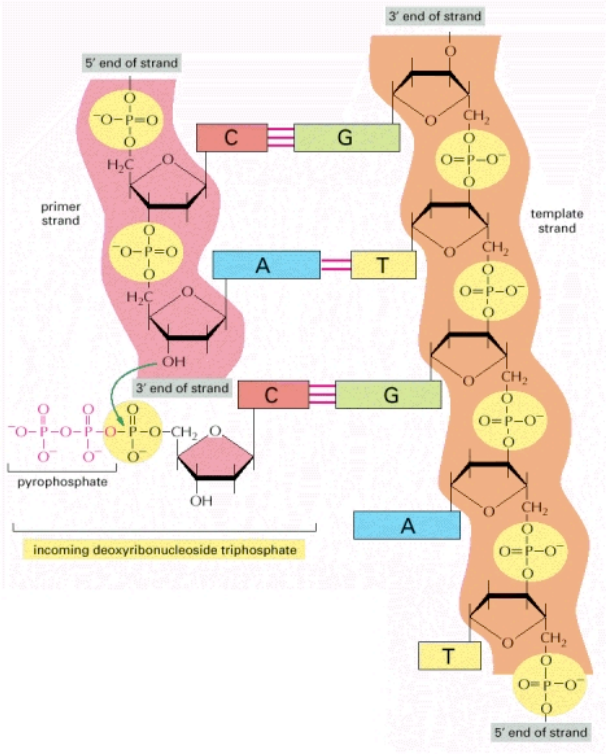



**Figure 3. Mechanism of polymerization.**
Cartoon depicting the main chemical step of the polymerization reaction. Template strand is colored in orange, primer strand is colored in pink, as well as the incoming nucleotide.

template, DNA with small gaps, DNA lesions or modifications). Finally, one of the key differences is the fidelity of synthesis of each DNA polymerase (Kunkel, 2004). The high fidelity of replicative polymerases at nucleotide insertion is reinforced by a proofreading 3'→5' exonuclease activity (Fig. 1). On the other hand, there are also unconventional DNA polymerases, as TdT, with a propensity to introduce mutations when required, i.e. to generate variability at antigen receptor genes. Currently a high number of DNA polymerases have been described, with representatives in virtually all kingdoms (reviewed in Goodman and Tippin, 2000). Classification in families was made based on similarities at the primary structure of their catalytic subunits (Braithwaite and Ito, 1993; Ito and Braithwaite, 1991). Table 1 shows the main members of the different families of DNA polymerases present in higher eukaryotes, and the 3D structure of one member of each family is shown in figure 2.

## 2. Polymerization Mechanism

DNA polymerases share a common mechanism. From the chemical point of view, these are the enzymes responsible for catalyzing the nucleophilic attack of the 3'-hydroxyl group of deoxyribonucleoside monophosphate (dNMP) residue located at the 3' position of the growing chain, known as primer-terminus, on the α-phosphorus of the deoxyribonucleoside triphosphate (dNTP) that will be incorporated. Most DNA polymerases require the presence of a DNA template strand that directs the synthesis of the growing chain according to the rules of base pairing defined by Watson and Crick, thus dictating the order of incorporation of dNTPs (Fig. 3). Therefore, polymerization always requires the

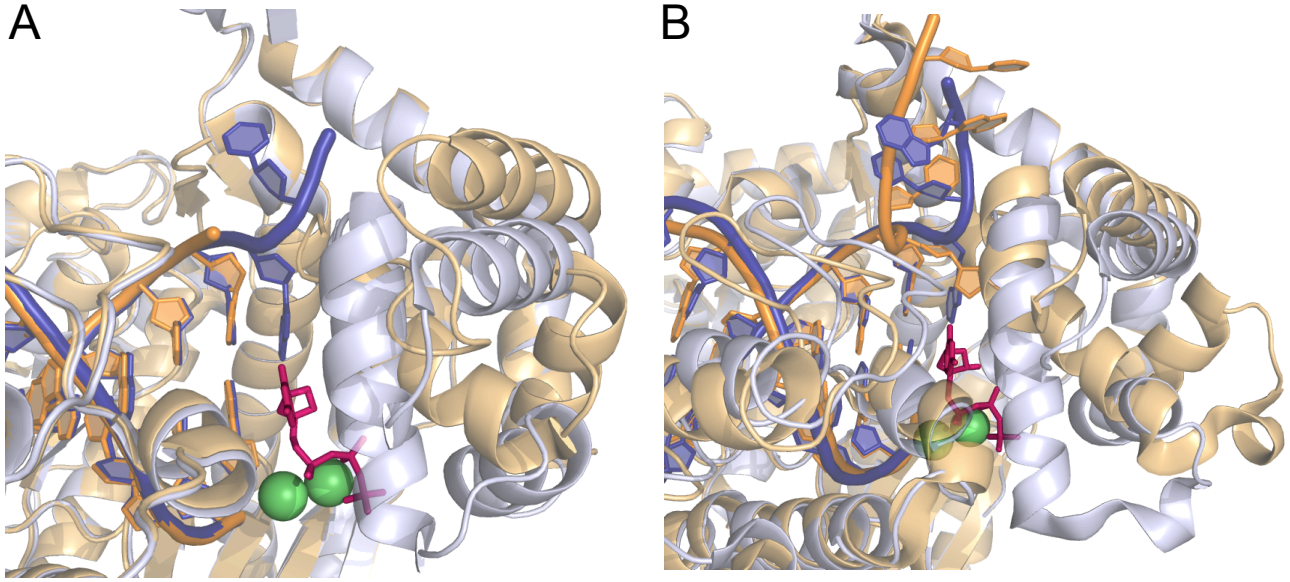



**Figure 4. Conformational change upon ternary complex formation.**
Superimposition of the binary (orange) and ternary (blue) complexes of T7 (A) or *Bacillus subtillis* (B) polymerases, showing the conformation change that the fingers domain undergoes during the binding of the incoming nucleotide (dark pink). The activating metal ions are shown in green. See the main text for details.

existence of a primer end, to provide a 3'-hydroxyl group that can be elongated to form the nucleotide chain of newly synthesized DNA. The molecule that acts as a primer can be a segment of DNA or RNA, and some DNA polymerases use the hydroxyl group of a specific amino acid (Ser, Thr or Tyr) of a terminal protein (Blanco and Salas, 1984; Salas, 1991).

Most enzymes exhibit large structural changes upon binding to the DNA substrate, and it is apparent that some of the substrate binding energy is used to organize the active site and orienting this substrate for catalysis. Similarly, the binding of a nucleotide substrate to a high-fidelity DNA polymerase induces a more subtle change in structure of the enzyme from an *open* state in the absence of nucleotide to a *closed* state after binding nucleotide (Fig. 4; reviewed in Johnson, 2010. It appears obvious that one role of the substrate-induced conformational change is to allow the rapid binding of substrates (and release of products) in the *open* state while affording optimal alignment of catalytic residues surrounding the substrate to promote catalysis in the *closed* state. The shape of the base pair formed determines the fate of the weakly bound nucleotide during the conformational change step. Only a correct base pair induces a structural change in which the enzyme closes around the base pair to form a tight, catalytic complex. If a mismatched nucleotide is bound, the enzyme does not close, but rather proceeds to a structure which promotes nucleotide release while reducing the rate of catalysis. Structural studies as well as sequence comparisons among polymerases strongly suggest the hypothesis that the phosphoryl transfer reaction of all polymerases is catalyzed by a two-metal ion mechanism (Fig. 5) originally proposed by analogy to the well studied two-metal ion mechanism in the 3′-exonuclease reaction (Brautigam and Steitz, 1998; Freemont et al., 1988; Beese and Steitz, 1991; Derbyshire et al., 1991). The first observation of a polymerase complex with both primer-template DNA and dNTP·$Mg^{2+}$ bound to the polymerase active site that directly showed the structural basis of a two-metal ion mechanism was a complex with rat Polß (Pelletier et al., 1994). These two-magnesium ions are bound by three carboxylates contained in the palm domain. A higher resolution structure (2.1 Å) of human pol β complexed with a gapped DNA substrate and dideoxy-CTP shows precise details of the interaction of two hexacoordinated, partially hydrated $Mg^{2+}$ ions interacting with the three phosphates (Sawaya et al., 1997). Metal ion A interacts with the 3′-hydroxyl of the primer strand and is proposed (Beese and Steitz, 1991) to lower the p$K_a$ of this group, facilitating its attack on the α-phosphate of the incoming dNTP. Metal ions A and B are also proposed to stabilize both the structure and charge of the pentacovalent transition state that occurs during the course of this reaction. Finally, metal ion B binds to and is proposed to facilitate the leaving of the β- and γ-phosphates. Chemically similar mechanisms of two-metal ion-catalyzed phosphoryl transfer reactions are used by many enzymes including ribozymes (Steitz and Steitz, 1993).

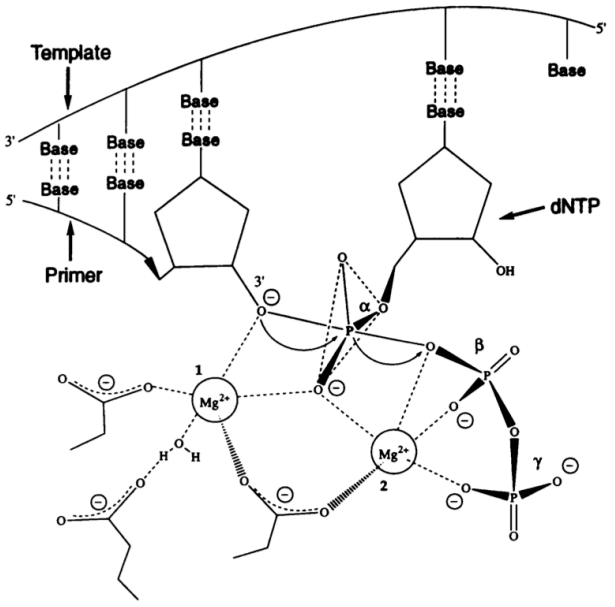



**Figure 5. General mechanism of nucleotidyl transfer.**
The active site features two metal ions that stabilize the resulting pentacoordinated transition state. Metal ion A activates the primer's 3'-OH for attack on the α-phosphate of the dNTP. Metal ion B plays the dual role of stabilzing the negative charge that builds up on the leaving oxygen, and chelating the ß- and γ-phosphates.

As reviewed by Johnson, 2010, the two metal ions alone are not sufficient for optimal specificity and efficiency of DNA replication. Mutagenesis has shown that residues in the fingers domain that contact the incoming nucleotide and templating base are critical for catalysis and fidelity (Astatke et al., 1998; Lam et al., 1998; Thompson et al., 2002). These residues may be distant from the incoming base during the initial binding of the nucleotide to the E:DNA *open* complex. The conformational change, which is triggered by the binding of a correct base

pair, brings these residues into contact with the incoming dNTP and templating base to align the reactants and facilitate catalysis by the two metal ions. Thus, one can think of the two metal ions as forming the core catalytic center, but movements in the recognition domain provide the basis for selective activity by bringing a correct incoming dNTP into proper alignment with the primer 3′OH for reaction and by providing additional catalytic residues.

Following the chemistry step, the enzyme must release pyrophosphate (PPi) and then translocate to allow the binding of the next nucleotide. Very little is known about these steps because all evidence suggests that both reactions are normally much faster than chemistry (Patel et al., 1991; Hanes and Johnson, 2008; Hanes et al., 2007). The most reasonable model for translocation is based upon a fast diffusion of the DNA between the nucleotide (N) and primer (P) sites. When the DNA product is still bound at the N-site (pre-translocation state), PPi can bind to reverse the chemical reaction by the process of pyrophosphorolysis producing dNTP. When DNA is already at the P-site (post-translocation state), it can bind the next nucleotide leading to primer extension. As the DNA rapidly diffuses between the N- and P-sites, the binding of dNTP captures the DNA at theP-site and the conformational change then locks the nucleotide and DNA in the *closed* state. Alternatively, PPi can trap the DNA at the N-site (Gotte, 2006). Translocation is usually fast and not kinetically significant.

**2.1 Replicative DNA Polymerases: B Family**

The family B polymerases are involved in nuclear DNA replication, except for the Polζ (Nelson et al., 1996). In eukaryotic cells, DNA replication is performed by at least three polymerases: Polα, Polδ and Polε (reviewed in Bebenek and Kunkel, 2004). The Polα complex has an associated primase activity, capable of synthesizing *de novo* the short RNA fragments used as primers for the synthesis of DNA strands during replication. Polα is able to extend these RNA primers, generating a short DNA primer, after which the elongation process is continued by Polδ or Polε (reviewed in Bebenek and Kunkel, 2004). These two enzymes are highly processive, a basic characteristic of replicative DNA polymerases that allows them to incorporate thousands of nucleotides without dissociating from DNA. This behavior is due to their ability to associate with the processivity factor (PCNA), which keeps them bound to the DNA (reviewed in Kelman, 1997). They also have a high nucleotide selectivity at the active site, resulting in high-fidelity synthesis. This is because the active site of replicative polymerases is adapted to accept Watson and Crick base-pairs, and reject those that do not adopt this geometry. In addition, the establishment of correct base pair promotes a series of conformational changes in the active site that do not occur with the entry of incorrect nucleotides, and facilitate the polymerization reaction (reviewed in Kunkel and Bebenek, 2000). Another important feature of these polymerases is their evolutionarily conserved 3'➔5' exonuclease domain (Bernad et al., 1989) contributing to proofread the incorrectly incorporated nucleotides, thus helping to increase fidelity (reviewed in Kunkel, 1988). Thus, these enzymes are suitable for faithful replication of large eukaryotic genomes.

**2.2 Mitochondrial DNA Replication: A Family**

Within this family the most prominent member is Polγ (Fig. 3). This enzyme is responsible for mitochondrial genome replication, and is the only known cellular DNA polymerase present in mitochondria (Kaguni, 2004). Like the replicative polymerases described above, its basic characteristics are its high fidelity of synthesis, high processivity and 3'➔5' exonuclease activity (reviewed in Bebenek and Kunkel, 2004). Recently two new members in this family, Polθ (Seki et al., 2003; Sharief et al., 1999) and Polv (Marini et al., 2003) have been identified and implicated in DNA repair processes.

**3. Genome damage and stability**

Survival and perpetuation of living beings depends on the maintenance and conservation of their genetic information. The integrity of the genome is constantly threatened by harmful agents from environmental sources (radiation, chemicals) and also by endogenous causes (products of its own metabolism). The combination of these genotoxic agents with the inherent instability of DNA can lead to an alteration of the chemical structure of DNA and

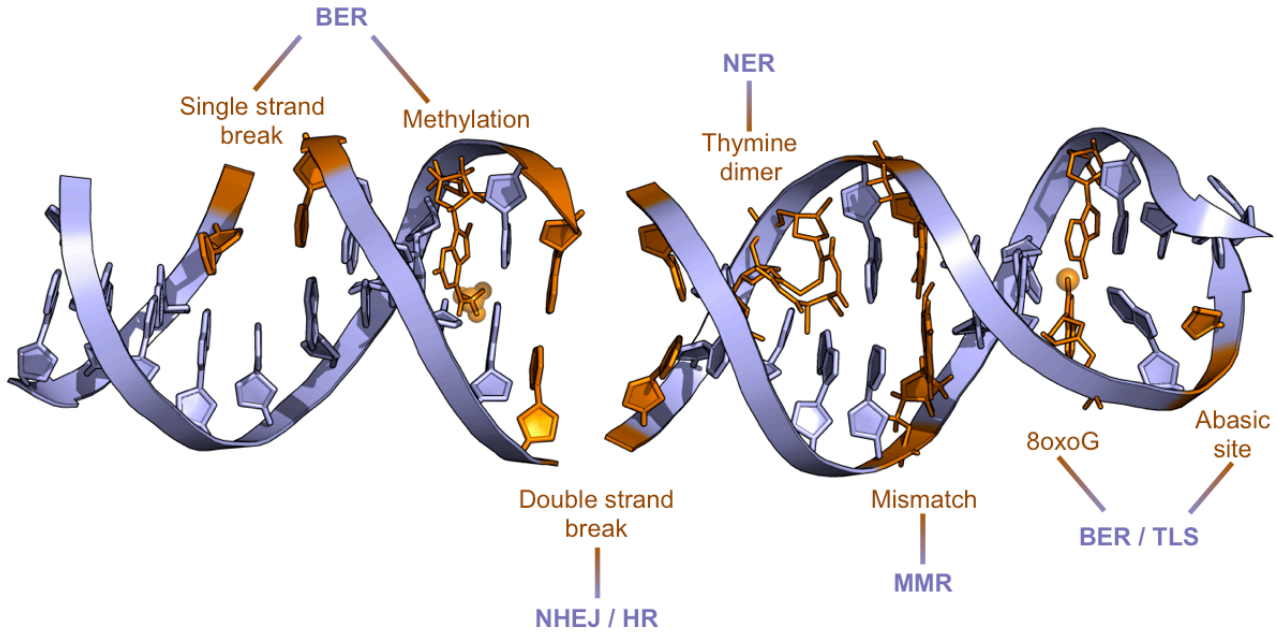



**Figure 6. DNA lesions caused by different damaging agents and repair and tolerance pathways.**

therefore of the encoded message (Hoeijmakers, 2001). The amount and diversity of injuries affecting the DNA is immense (Fig. 6); it is estimated that endogenous damage alone causes about 20.000 injuries per day in a cell (Lindahl et al., 1993). In general, three types of agents cause DNA damage. Firstly, environmental agents such as ultraviolet light (UV), ionizing radiation (IR) and many chemicals that can cause alterations in the structure of DNA. Secondly, products generated during normal cellular metabolism, such as reactive oxygen species (ROS), which can produce up to 100 distinct modifications in DNA (Cadet et al., 1997). Thirdly, some chemical groups in the DNA can be modified spontaneously under physiological conditions, such as the hydrolysis of nucleotides, generating abasic sites or deamination of bases which can lead to changes in the sequence of DNA (i.e., deamination of the base cytosine causes the base uracil). Eukaryotic cells have a network of signals which allows them to respond to different types of damage, specifically activating one or more repair mechanisms, cell cycle arrest, inhibition of replication and DNA transcription and even cell death by apoptosis (Huen and Chen, 2008).

Therefore, to maintain the integrity of the genome, DNA damage must be eliminated. Thus, the processes of cell replication and transcription can be carried out faithfully and efficiently. Living beings have different highly conserved repair mechanisms which are generally specific to each type of damage. At least there are four main DNA repair pathways: Base Excision Repair (BER), Nucleotide Excision Repair (NER), Mis-Match Repair (MMR) and Double Strand Break Repair (DSBR). However, many lesions in DNA bases escape repair mechanisms and block the progression of transcription and replication. For this reason, organisms have developed a mechanism called Trans-Lesion Synthesis (TLS), by means of specialized DNA polymerases able to insert nucleotides opposite the damage and extend it, allowing continued replication (Prakash et al., 2005).

## 4. DNA repair routes involving DNA polymerases

### 4.1 Nucleotide Excision Repair (NER)

The mechanism of NER operates primarily on bulky damage, which distorts the DNA double helix obstructing the normal processes of DNA replication and transcription. Two subtypes of NER have been stablished, regarding the substrate acted upon: one is activated when the lesion occurs anywhere in the genome (Global Genome NER, GG-NER) and the other acts on actively transcribed regions, using the transcribing RNA polymerase as a damage sensor, stopping at the lesion, hence its designation as transcription coupled repair, TCR. The difference between the two mechanisms is basically the protein complexes that recognize lesions (de Laat et al., 1999). The rest of the repair process is common: the

damaged DNA double helix is opened by a number of helicases that are part of the TFIIH complex (namely XPB and XPD). The bubble generated about 30 base pairs around the lesion is recognized by a protein binding to single-stranded DNA, the replicative protein A (RPA), which specifically binds to the undamaged strand. Thus, the damaged strand can be recognized by the endonucleases XPG and ERCC1/XPF, producing cuts on either side of the lesion, followed by the removal of a fragment of single-stranded DNA between 24-32 bases containing the damaged base. Finally, the gap is filled with high fidelity by the replication machinery.

### 4.2 Mis-Match Repair (MMR)

The mechanism of MMR removes nucleotides misincorporated as a result of slippage during replication of repetitive DNA sequences or during recombination. This mechanism is essential to avoid mutagenesis in the DNA, and involves numerous enzymatic activities, including some of those involved in NER (Schofield and Hsieh, 2003). The repair process begins with the recognition of the injury by heterodimeric protein complex MSH2/6 or MSH2/3. Once the damage is located, Mlh1/Pms1 or Mlh1/Pms2 protein complexes are recruited and in turn interact with replicative factors. The system is thus informed of which DNA strand has to be repaired: the newly synthesized. The process of nucleotide excision to remove the mismatch and the subsequent DNA resynthesis involves many proteins, including DNA polymerases Polδ and Polε, RPA, PCNA or endonuclease FEN-1 (Schofield and Hsieh, 2003).

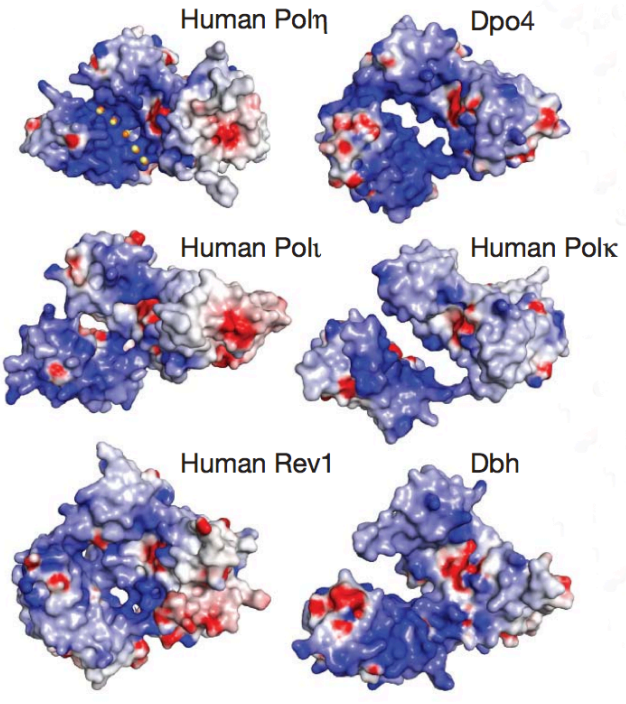



**Figure 7. Portrait of the Y family polymerases**
The DNA binding surfaces of various Y family polymerases shown with electrostatic surface potentials. Blue and red represent the positive and negative charge potential, respectively. The thumb domain in all cases and the N-clasp of Polκ are removed for clarity. Five phosphorus atoms at position 21 to 25 in the human Polη Nrm (yellow) and TT4 (orange) structures are also shown. Modified from Biertümpfel *et al*, 2010.

### 4.3 Trans-Lesion Synthesis (TLS)

The occurrence of different lesions in DNA during the S phase of cell cycle, that cannot be repaired before they are encountered by the replication machinery, requires the existence of mechanisms to resolve the block in the replication fork caused by unrepaired lesions that change the normal geometry of the double helix (abasic sites, contiguous covalent bonds between bases, etc.). The disturbance in the template strand causes the shutdown of the replicative DNA polymerases, Polε and Polδ, due to their strict requirement for a correct geometry in base pairing. To overcome this barrier, a number of DNA polymerases, copying/coping with a damaged template, have evolved. These enzymes alleviate the problem of replication blockage, but create the need for a later repair process to eliminate the original lesion. Therefore, these polymerases are not repair enzymes, but enzymes involved in damage tolerance that allow cell survival in the presence of DNA damage. These polymerases are grouped in the Y family, except Polζ that belongs to the B family. DNA polymerases included in the Y family are Polη, Polι, Polκ and REV1 (Fig. 7; Prakash et al., 2005).

#### *4.3.1 Y family*

The members of the Y family have a number of common characteristics: they all are induced in the presence of agents causing DNA damage, their mode of synthesis is distributive, and their catalytic efficiency and nucleotide insertion fidelity are both low (Kunkel, 2004). Moreover, although all these polymerases share five conserved sequence motifs at their N-terminal part, the C-terminus is different in each of them, and usually contains nuclear localization motifs and domains devoted to

interaction with PCNA and other polymerases. Each Y polymerase is specialized in allowing synthesis through a particular type of lesion, although they have certain common characteristics: they lack the proofreading 3'➔5' exonuclease activity, typical of the replicative DNA polymerases, and differ further at the structural level, since Y family polymerases have *thumb* and *fingers* subdomains smaller than those of the replicative enzymes. This translates into an active site open and accessible at the *palm* subdomain, which can accept different types of lesions and incorrect base pairs, and implies a low-fidelity synthesis (reviewed in Yang, 2005). All Y family members interact with PCNA, which has led to different models to explain the possible trafficking of proteins (Pages and Fuchs, 2002) and its post-translational regulation by ubiquitination and sumoylation (Hoege et al., 2002; Ulrich, 2004).

### *4.3.2 Substrate binding & polymerization mechanism*

Four structural subdomains are found in each polymerase domain. The first 250–350 residues including the five signature motifs (Kulaeva et al., 1996) constitute the catalytic core of the polymerases and form the thumb, palm, and finger subdomains as found in all known DNA and RNA polymerases. Despite a lack of apparent homology, the palm domain of the A, B, and Y family polymerases, as well as reverse transcriptases, are highly conserved, and the three carboxylates essential for the catalysis are located on identical structural elements (Yang, 2005). Although the secondary structures vary broadly in the thumb and finger domains across the different polymerase families, their location in the tertiary structure and roles in interaction with DNA and nucleotide substrate are conserved among all polymerases. However, the thumb and fingers are distinctively smaller in Y family polymerases.

At the C terminus of the catalytic core, 100 residues form a structurally conserved domain unique to the Y family polymerases. It has been called little finger (LF) domain, based on the analogy to a right hand (in addition to palm, thumb, and finger) and its role in DNA binding (Ling et al., 2001). In contrast to replicases, whose finger domain interacts with the replicating base pair and undergoes the largest conformational changes upon substrate binding, the most mobile domains in Y family polymerases are the LF and thumb, that sandwhich the upstream DNA. In the absence of DNA substrate, the LF can be closely associated with the catalytic core through a tether as observed in Dbh (Silvian et al., 2001), or wildly flexible as in polκ (Uljon et al., 2004). The mobility of the LF and thumb can alter the positions of DNA substrate relative to the catalytic core and, consequently, the activity of the polymerase (Ling et al., 2003; Rechkoblit et al., 2006).

The active site of Y family polymerases is preformed before substrate binding, in contrast to that of the A, B, and RT family polymerases (Zhou et al., 2001; Yang, 2003). Because of the small finger and thumb, the active site is also highly solvent-exposed and not as geometrically constrained to reject non-Watson–Crick base pairs. The ''spaciousness'' of the active site gives grounds for erroneous base pairing and the ability to accommodate bulky DNA lesions (Ling et al., 2001). As for the specific substrate binding, human Polη acts like a molecular splint to keep the damaged DNA straight and rigid owing to a continuous and highly positively charged DNA-binding surface that interacts extensively with the four template nucleotides immediately upstream of the active site (Biertumpfel et al., 2010; Fig. 7). A ß-strand in the LF domain is nearly parallel to the template strand, and every other main-chain amide donates a hydrogen bond to the template phosphate. In contrast, the template-binding surface in other Y family polymerases has a gap or holes owing to separations between LF and the catalytic core (Fig. 7). The gap is particularly large in Polk, and the divided protein requires an N-terminal 33 appendage (N-clasp) for stabilization (Lone et al., 2007). Polk and Dpo4 probably use the structural gaps to accommodate bulky minor-groove adducts during TLS template looping out as a means of lesion bypass (Ling et al., 2004; Wilson and Pata, 2008).

For a conventional polymerase reaction, the 3'-OH of a primer strand and the α-phosphate of a dNTP have to be placed adjacent to each other and oriented for the phosphoryl transfer reaction. In high-fidelity polymerases, when a correct dNTP approaches a template–primer duplex, the finger domain undergoes a large conformational change, and the active site becomes "closed" in this structure, the 3'-OH and alpha-phosphorus of dNTP are then aligned with the catalytically essential metal ions and carboxylates for the chemical bond

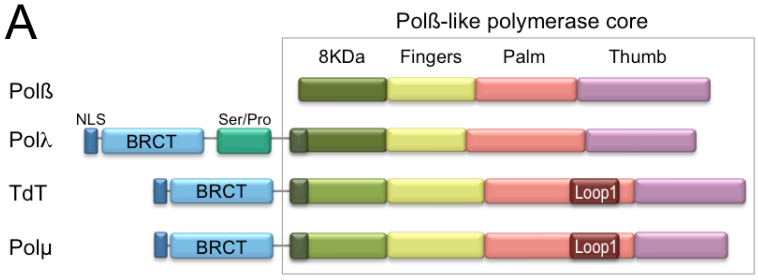
formation. The fidelity is thus achieved mainly in two steps: 1) a large conformational change of the finger domain; and 2) alignment of reactants and catalysts in the active site. Another peculiarity of the Y family polymerases is the lack of a definitive conformational change in the finger domain in response to binding the incoming nucleotide that forms a Watson-Crick base pair with the template base. The lack of an "induced-fit" conformational change surrounding the active site seems to remove another barrier to the ability to make mismatched base pairs by Y family polymerases, resulting in low fidelity and translesion synthesis. Interestingly, aside from the three catalytic carboxylates chelating the divalent metal ions and the residues interacting with the triphosphate moiety of an incoming nucleotide, protein residues interacting with the template base and incoming nucleotide are not conserved among the Y family polymerases. It is hypothesized that the variation of residues in contact with the replicating base pair allows each polymerase to attain its selection of a specific lesion to bypass and a favorite nucleotide to incorporate (reviewed in Yang and Woodgate, 2007).

### 4.4 Base Excision Repair (BER)

This repair pathway corrects DNA lesions affecting the DNA bases (i.e., presence of uracil in DNA modification or loss spontaneous bases). Two BER pathways in mammals have been described (reviewed in Dogliotti et al., 2001): a "long patch" BER, which involves the replacement of 2 to 10 nt, and a much more widely used "short patch" BER that includes replacing a single nucleotide, the one primarily involving Polß. In this process, the glycosidic bond between the damaged base and deoxyribose is removed, either spontaneously or by means of a DNA glycosylase, to generate an abasic site in dsDNA. Then, the phosphodiester bond in position 5' of the abasic residue is eliminated by a specific endonuclease (APE), after which Polß is involved, carrying out a polymerization reaction that replaces the damaged nucleotide, and eliminates the ribose-phosphate residue by the dRP lyase activity present in the 8 kDa domain. Finally, the reaction is completed by the action of a DNA ligase. Polß involvement in this mechanism has been demonstrated both *in vitro* (Kubota et al., 1996) and *in vivo* (Sobol et al., 1996). In addition to its enzymatic activities, certain proteins involved in BER are known to interact Polß, such as the endonuclease APE (Bennett et al., 1997), Ligase 1 (Prasad et al., 1996; Dimitriadis et al., 1998) and the heterodimer XRCC1/Ligase 3 (Caldecott et al., 1996; Kubota et al., 1996), which also may be essential for this process (Wilson and Kunkel, 2000). On the other hand, several studies have suggested the involvement of other DNA polymerases in BER when Polβ is absent or during repair of certain types of oxidative damage. One of them is another member of the X family, Polλ, which also contains a dRP lyase activity and has the same preference for DNA substrates with small gaps and a 5'-P group (Garcia-Diaz et al., 2000; Garcia-Diaz et al., 2001). Polι, a family Y polymerase, also has DRP lyase activity, and may participate in a more specialized BER (Bebenek et al., 2001). Polγ also displays dRP lyase, being the only DNA polymerase present in mitochondria (Longley et al., 1998). For this reason Polγ would be responsible for all DNA synthesis in this organelle, including the processes of repair of DNA damage.

#### *4.4.1 Polß: a paradigm of X family polymerases*

Polß is the smallest polymerase known in mammals (39kDa) and has enzymatic activities that are specifically appropriate for the filling-in of small gaps. As the first X family polymerase crystallized, Polß served as a model for the study of the other members of the same family. All of them share a common structural organization of the Polß basic core, consisting of an N-terminal domain (8 kDa), connected to the polymerization domain (31 kDa) by a protease-sensitive region (Prasad et al., 1998). The polymerization domain is divided into three subdomains, named fingers, palm and thumb. Due to its "left-handed" nature, the thumb and fingers sub-domains of Polß are functionally equivalent to the fingers and thumb sub-domains of the "right-handed" replicative polymerases, respectively (Beard and Wilson, 2000). Aside from this common basic structure, individual members of this family may have additional domains (Fig. 9) that give them certain special features. Polß is undoubtedly the most studied member of the X family. Since its discovery numerous biochemical and structural studies have been conducted, which enabled the association of Polß with a number of functions.



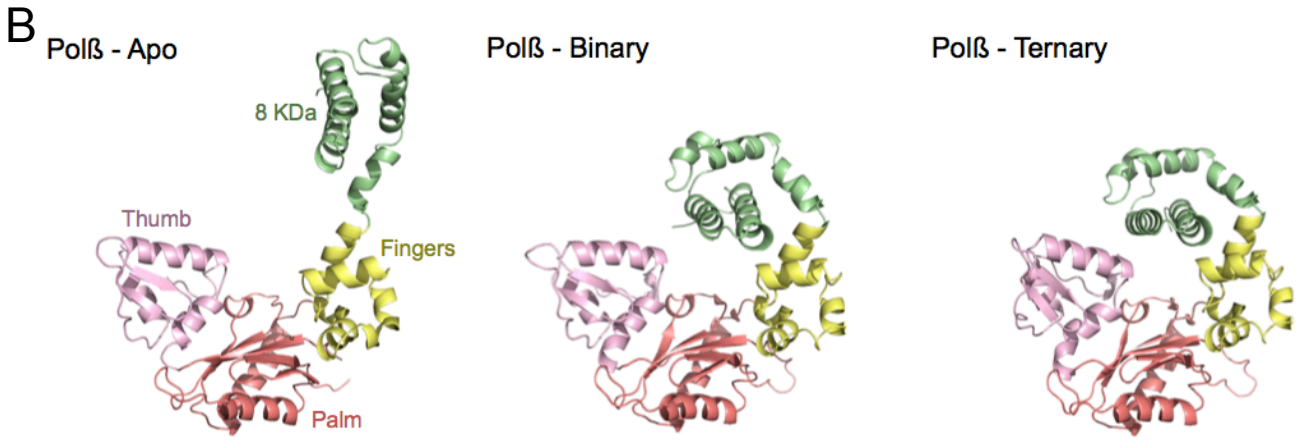




**Figure 8. Structural organization of the human X family members.**
A) On the basic Polß-like domain scheme (8 kDa, fingers, palm and thumb) are shown the additional elements present in the other members (Ser/Pro rich region, nuclear localization sigal -NLS-, BRCT domain). B) Crystal structures of the three Polß conformations (Apoenzyme, E:DNA Binary complex and E:DNA:dNTP ternary complex). Domain colouring as follows: 8 kDa in green, fingers in yellow, palm in red, thumb in pink.

(reviewed by Beard and Wilson, 2000 and amended by Idriss et al., 2002).

### *4.4.2 Enzymatic features of Polß*

The polymerization activity is known be carried out by the catalytic domain (31 kDa), and follows a kinetic mechanism similar to that of most DNA polymerases, using two metal ions (Steitz et al., 1994). After the interaction with DNA (Tanabe et al., 1979), Polß uses a two-step mechanism for binding the nucleotide (Ahn et al., 1997), implying a conformational change consistent with a mechanism of "induced-fit" (Sawaya et al., 1997). One of the main features of Polß is that, unlike the replicative polymerases, lacks the 3'➔5' exonuclease activity (Tanabe et al., 1979), which implicates a decrease in fidelity during DNA synthesis. It is known that *in vitro* Polß is error-prone, with a frequency of base substitution errors of $10^{-3}$-$10^{-4}$ (Ahn et al., 1997; Kunkel, 1985; Osheroff et al., 1999). This is further supported by the fact that Polß is a distributive enzyme, which would tend to generate sliding-mediated deletions (Kunkel, 1985; Osheroff et al., 1999; Werneburg et al., 1996). However, it has been suggested that the effective fidelity of the polymerase may increase in vivo the presence of other enzymes endowed with 3'➔5' exonuclease activity, which may eliminate the Polß inserted errors (Shevelev and Hubscher, 2002). It is also known that the activity and fidelity of this polymerase is greatly enhanced when acting on DNA gaps of small size, its preferred substrate. Based on these characteristics, it is an appropriate enzyme for DNA repair mechanisms.

### *4.4.3 Polß structural organization*

X-ray crystallography studies have unveiled the structures of the Polß apoenzyme or in complex with different DNA substrates, and the conformational changes that result as a consequence of the binding of nucleotide. This has contributed greatly to elucidate functional aspects of this polymerase. The structure of free Polß showed that the polymerization domain (31 kDa), composed of subdomains fingers, palm and thumb, adopts a structure in space similar to a partially closed left hand, similar to other DNA polymerases not evolutionarily related (Brautigam and Steitz, 1998; Holm and Sander, 1995). Its N-terminal domain (8 kDa) is shown to be located away from the polymerization domain in this structure (Fig. 8, Polß - Apo).

The crystallization of Polß complexed with different DNA molecules allowed to establish the structural basis of substrate preference of this enzyme for small DNA gaps. The Polß complex with a template-

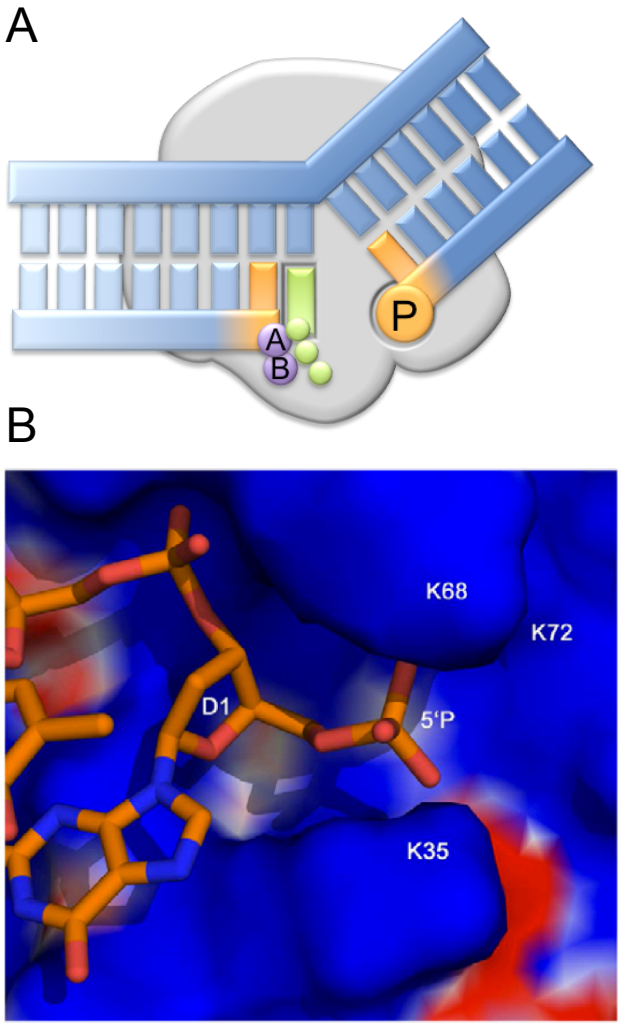



**Figure 9. 5'-P binding pocket.**
A) Cartoon depicting Polß (gray), the incoming dNTP (green), the two activating metal ions (A and B, purple) and the DNA substrate (blue and orange). The 5'-P group is shown as an orange circle. The 3'-end of the primer strand is shown in orange. B) Electrostatic surface of the positively charged 5'-P binding pocket in Polß. The DNA substrate is shown in sticks.

primer substrate shows an arrangement in which the 8 kDa domain is not interacting with the DNA, being positioned at a distance from the polymerization domain (Beard and Wilson, 1998; Pelletier et al., 1994). In contrast, the crystal structure of Polß complexed with a 1-nt gapped substrate shows that the 8 kDa domain binds to the 5'-phosphate of the gap, while also interacting with the substrate through the thumb subdomain (Beard and Wilson, 1998; Sawaya et al., 1997), thus promoting a more stable binding of polymerase to this substrate (Fig. 9).
The resolution of DNA structures including also the incoming nucleotide has allowed to further study the polymerization mechanism and conformational changes that occur during catalysis. It has been determined that the entry of the nucleotide causes the formation of a closed complex by induction of a conformational change of the thumb subdomain (Sawaya et al., 1997), which allows it to establish a set of fundamental interactions for nucleotidyl transferase reaction, having also a decisive influence on the fidelity of the reaction (Beard et al., 2004; Kunkel and Bebenek, 2000). In addition, the interactions stablished with the primer through a helix-hairpin-helix (HhH) motif located in the fingers domain (Doherty et al., 1996) helps to close the nucleotide binding site. It has also been proposed that after the incorporation of the nucleotide, a new conformational change takes place, regenerating the open complex, and possibly facilitating the removal of pyrophosphate after catalysis (Pelletier et al., 1996).

### *4.4.4 8 kDa domain*

It is known that the 8 kDa domain binds ssDNA (Kumar et al., 1990) and interacts strongly with the 5'-phosphate of a DNA gap (Fig. 9; Prasad et al., 1998), which influences their characteristics polymerization. While Polß is a distributive enzyme on template-primer substrates (Abbotts et al., 1988; Werneburg et al., 1996), this is not true for gapped substrates. If the gap is small (up to 6 nt), the enzyme is able to bind and carry out processive synthesis. Conversely, if the gap is bigger than 6 nt, Polß binds to its 5' end but does not perform any synthesis (Singhal and Wilson, 1993). This evidence, coupled with structural data, define a preference for small DNA gaps, resulting also in an increased fidelity of synthesis (Chagovetz et al., 1997; Singhal and Wilson, 1993). Only when a 3′-OH is "within reach" will it be extended (Fig. 10). In addition to defining the substrate specificity, the 8 kDa domain of Polß determines the principal function proposed for this enzyme, since it contains its dRP lyase activity. This activity, originally described in *E. coli*, (Franklin and Lindahl, 1988), involves the removal of the abasic residue from the 5' end of DNA, crucial for the BER mechanism.

### *4.4.5 Polymerization mechanism*

As explained, Polß prefers to bind gapped DNA substrates that have a 5′-phosphate on the downstream strand. Interactions between the 8-kDa domain and the thumb domain are altered upon binding the correct nucleoside triphosphate. In the crystal structure of Polß bound to ddCTP and a one-

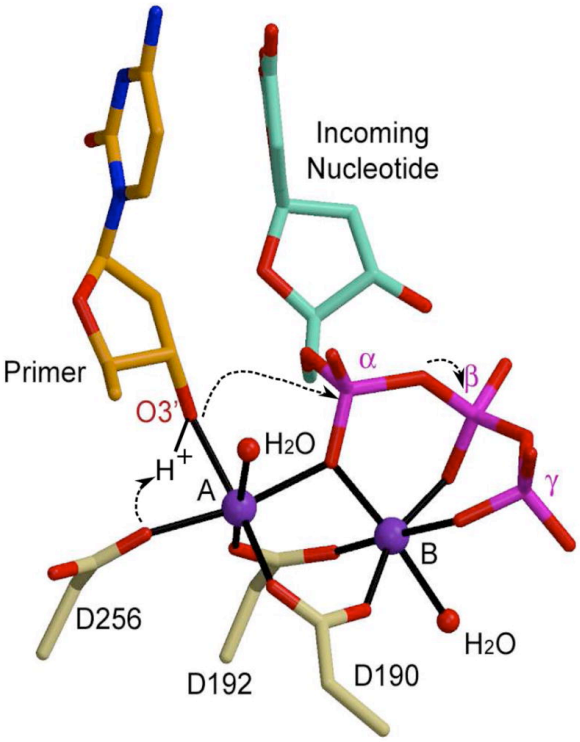



**Figure 10. Catalytic mechanism for nucleotidyl transfer.**
The primer terminal residue (orange) is drawn in stick, with the O3′ in red. The incoming nucleotide is cyan, with its phosphates in magenta. Protein residues modeled from the ternary structure of Polß (PDB code 2FMS) are khaki. The bridging nitrogen from the dUMPNPP has been modeled as an oxygen. Magnesium ions are purple and water molecules completing coordination of the metal ions are small red spheres. Modified from Moon *et al*, 2007.

nucleotide DNA gap, the DNA is bent about 90º (Fig. 9). Accordingly, the DNA template does not travel through the hole in the doughnut-like structure, but instead moves across the surface of the HhH. The sharp bend occurs at the 5'-Phosphodiester bond of the templating base in the one-nucleotide gap and is also observed in the product complex where Polß is bound to nicked DNA (Sawaya et al., 1997). The dNTP binding pocket of DNA polymerases is formed by the template base, DNA duplex terminus, and enzyme. In the ternary substrate complex of Polß, the nascent base pair (templating and incoming nucleotides) is sandwiched between the duplex DNA terminus and R-helix N. It is observed that normal Watson-Crick pairing, as well as aberrant primer termini, has a strong influence on correct nucleotide insertion (Beard et al., 2004).

In general, Polß utilizes a similar kinetic mechanism as most other DNA polymerases (Fig. 10). Steady-state kinetic analyses indicate that it follows an ordered addition of substrates (Tanabe et al., 1979). Employing pre-steady-state kinetics, it has been shown that Polß utilizes a two-step nucleotide binding mechanism (Ahn et al., 1997) similar to that determined for other DNA polymerases that have been examined in this manner. Initial binding of a nucleotide by the enzyme places the nucleotide within the active site. A subsequent conformational change results in the alignment of the catalytic groups and rapid chemistry. Binding of the correct nucleotide facilitates this conformational change and leads to tighter binding, whereas binding of the incorrect nucleotide hinders the conformational change resulting in weak binding. This "induced-fit" mechanism is consistent with the numerous structural changes that occur upon binding of a correct nucleotide (Sawaya et al., 1997). The templating base must be precisely positioned for the polymerase to faithfully check correct geometry. Comparing the position of the templating base in the open and complexes indicates that the template base changes position to base pair with the incoming nucleotide (Beard et al., 1996; Sawaya et al., 1997).

Polß is, thus, well suited for filling short gaps during DNA repair because it simultaneously binds both the 5′ and 3′ ends of short gaps. DNA binding and gap filling are well characterized for 1nt gaps, and Polß processively and systematically fills short gaps in which the 5′ end of the gap contains a 5′-phosphate, but the location of yet-to-be-copied template nucleotides in these longer gaps is unknown.

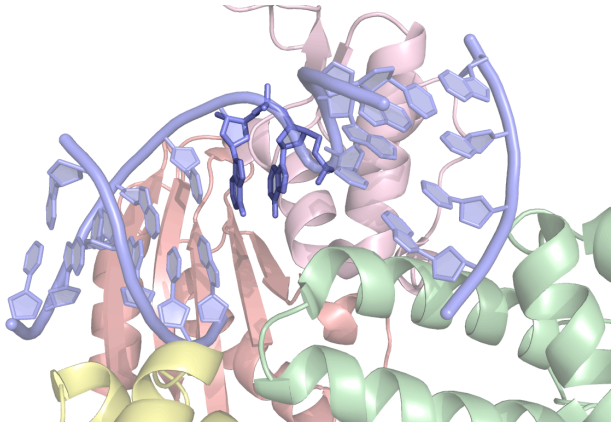


**Figure 11. Polλ complexed with a 2-nt gap.**
Crystal structure of Polλ with a 2-nt gap (blue) showing the erroneous positioning of the 3'-OH of the primer strand in the -1 position. The two templating bases are shown in dark blue. Domain colouring as follows: 8 kDa in green, fingers in yellow, palm in red, thumb in pink.

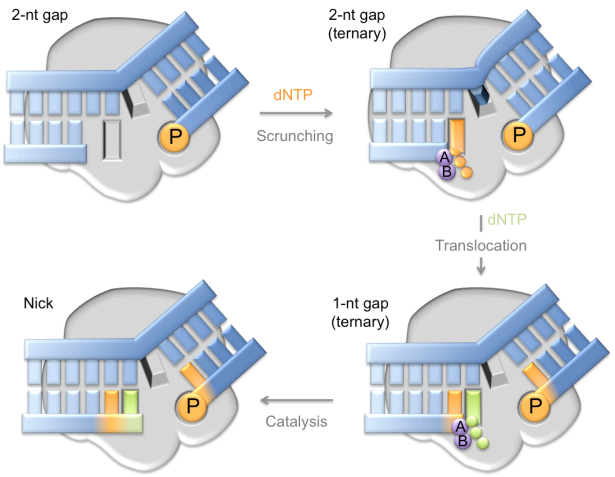



**Figure 12. Scrunching during gap filling.**
A simplified model of the steps during gap filling by Polλ in a gap longer than 1 nt. Each of the steps corresponds to a crystallized intermediate. The DNA polymerase is represented as a grey 3D surface, the DNA substrate is colored blue, the 5'-P is shown as an orange circle and the incoming nucleotie is depicted in green.

Moreover, in the first crystal structure of Polλ, another member of the X family, which included the polymerase in an inactive binary complex with a 2-nt gap, showed that the 8 kDa domain was bound to the 5'-phosphate end of the gap, but the polymerase active site was not productively engaged at the 3' end of the gap (Garcia-Diaz et al., 2004; Fig. 11). This structure suggested that, in the absence of an incoming nucleotide, the polymerase might preferentially bind to the 5' end of a gap, a feature that most surely applies to Polß as well. These observations make it necessary to invoke a conformational change leading to an intermediate where, either through DNA or protein rearrangements, the polymerase active site would bind the 3' DNA end in a catalytic conformation, whereas the 8 kDa domain would remain bound to the 5' end of the gap. These movements and fittings could be dependent on dNTP binding. In agreement with this hypothesis, a more recent structure (Garcia-Diaz et al., 2009) shows Polλ bound to DNA while attempting to incorporate a correct nucleotide opposite the first of the two single-stranded nucleotides in the gap, with a "scrunched" template strand, where the second templating nucleotide in the 2-nt gap is extrahelical (Fig. 12). Molecular dynamics simulations with longer gaps are consistent with the idea that Polλ can accommodate additional uncopied template nucleotides while maintaining a conformation similar to that observed in the crystal structures. The models further suggest that, when given multiple choices, the scrunching pocket is more likely to accommodate the nucleotide immediately adjacent to the templating nucleotide. The identification of this binding pocket for the second template nucleotide in the gap brings to three the number of DNA substrate–binding pockets observed in Polλ, the other pockets being for the nascent base pair and the phosphate at the 5′ end of the gap (Fig. 12). Polß is predicted to have the same scrunching pocket as Polλ due to the many structural and biochemical similarities of these polymerases, including a similar behaviour when filling in small gaps. Thus, from the currently available structures, and the similarity between Polλ and Polß, a common picture of gap filling emerges wherein correct dNTP binding induces several conformational changes in the polymerase and the DNA, especially in the template strand. These changes result in assembly of the binding pockets for the nascent base pair and the second nucleotide in the gap. Once the dNTP complementary to the first templating base of the gap is bound, catalysis can proceed and translocation occurs. This translocation brings the scrunched second templating base into the pocket of the nascent base pair and the complementary dNTP is then incorporated, and the function of the polymerase is fulfilled with the formation of a directly ligatable nick.

### 4.5 Double Strand Break Repair

The double-stranded breaks in DNA (DSB) are particularly harmful to the cell. DSBs may lead to loss of genetic information and cell death, or chromosomal rearrangements (translocations, inversions, deletions) that can lead to carcinogenesis (Mills et al., 2003). DSBs can be generated in a physiological manner during replication and meiosis (Michel et al., 1997; Shinohara and Ogawa, 1995). It is estimated that human cells undergo about 10 DSBs per cell cycle spontaneously, most of which occurs during replication when the replication fork encounters a single strand break (SSB) (Paques and Haber, 1999). Also, this type of breakage is caused by the action of endogenous chemicals, from the cellular metabolism (Karanjawala et al., 2002), or by the action of exogenous agents such as ionizing radiation (Morgan et al., 1996; Ward, 2000). In response to this type of damage the cells promote

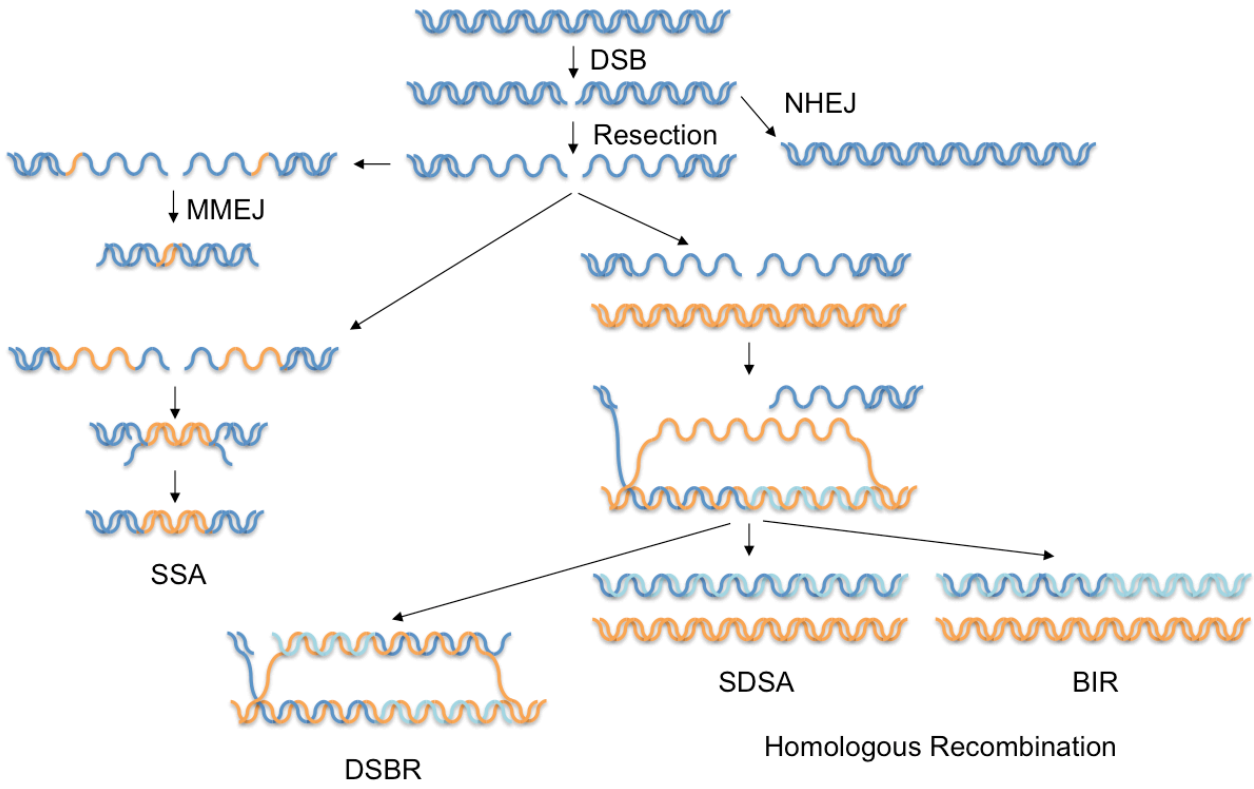



**Figure 13. The repair of DNA double-strand breaks (DSBs).**
DSBs can be repaired using several different mechanisms. Both ends can be simply rejoined with little or no further processing (nonhomologous end-joining; NHEJ) or can be repaired using homologous sequences (red DNA; homologous recombination) after 5′–3′ degradation has occurred (resection). The 3′-OH group exposed after resection can be used to prime DNA synthesis using a homologous region as a template after DNA strand invasion. The newly synthesized DNA (light blue) can then be joined with the 5′ end of the resected strand forming a double Holliday junction (double-strand break repair; DSBR), or can be displaced and reannealed (synthesis-dependent strand annealing; SDSA); or DNA synthesis can continue to the end of the chromosome (break-induced replication; BIR). If two homologous regions flank the DSB, they can anneal after being exposed by DNA resection (single-strand annealing; SSA), which causes the deletion of the intervening region. An additional mechanism that shares components with both SSA and NHEJ, and uses short homology stretches (usually 2–3 bp) flanking the DSB, can also be used (microhomology mediated end-joining; MMEJ). Modified from Huertas, 2010.

the efficient repair of damage to restore the integrity and functionality of the genome.

There are two main mechanisms of DSB repair (Fig. 13): homologous recombination (HR) and non homologous end-joining (NHEJ). HR is responsible for repairing the damage by using the genetic information of the sister chromatid or homologous chromosome, and thus is an error-free repair mechanism. Repair by NHEJ involves joining the two DNA ends using little or no homology, and this can lead to increased mutagenicity. In mammalian cells, the predominant mechanism for DSB repair is NHEJ, which operates throughout the cell cycle. On the other hand, HR occurs during late S or G2 phase, when the sister chromatid is close (Lieber, 2008). It should be noted that both mechanisms, HR and NHEJ may compete but also occur in a cooperative manner (Shrivastav et al., 2008).

### *4.5.1 Homologous recombination*

Homologous recombination (HR) is a ubiquitous cellular pathway that performs template-dependent, high-fidelity repair of complex DNA damage caused by endogenous and exogenous sources including DSBs and inter-strand crosslinks. Eukaryotic cells possess several mechanisms to repair DSBs by HR. There are at least three: Gene Conversion (GC; including Double Strand Break Repair –DSBR- and Synthesis Dependent Strand Annealing -SDSA), Break-Induced Recombination (BIR) and Single-Strand Annealing (SSA).

The DSB repair by HR requires the processing of DNA ends by a 5’➔3’ exonuclease that removes nucleotides from the 5’ end. This will generate large fragments of 3’-protruding ssDNA, a platform that allows binding of various proteins resulting in a nucleoprotein filament. RPA is responsible for removing DNA secondary structures allowing

binding of Rad51. Rad52 interacts directly with Rad51 and RPA, being the protein responsible for displacing RPA to allow binding of Rad51 to single-stranded DNA, becoming an essential protein in the process. Other proteins such as Rad55 and Rad57 also participate in the formation of the filament. The Rad51 nucleoprotein filament can interact with a second molecule of DNA, both double- or single-stranded, promoting the invasion of the sister chromatid. With the help of Rad54, chromatin remodeling and DNA unwinding are promoted, allowing exchange with homologous DNA.

*4.5.1.1 Gene conversion (GC)*

Gene conversion is defined as a transfer of non-reciprocal genetic information from one molecule to its counterpart. Usually occurs between two alleles of a gene. This mechanism is further subdivided in DSBR and SDSA. They differ on whether the products obtained contain crossovers or not (Paques and Haber, 1999; Symington, 2002).

In both cases the invasion of the chain occurs in the same manner (Fig. 13). Once generated the DSB, ends are equally processed by removing nucleotides in the 5'➔3' direction. Therefore, large single-stranded fragments are created and invade the homologous sequence. Thus, after displacement of one of the strands form a heteroduplex DNA, a D-loop structure is formed. Then the synthesis of DNA in the heteroduplex can occur, at the 3' end of the original site of rupture. Sister chromatid provides a template for an error-free repair to take place. In the DSBR model, an X-shaped structure called the Holliday junction (HJ) is formed, resulting from the invasion and subsequent cross-linking of sister chromatids. Its resolution creates either cross-overs or non cross-overs. By contrast, in the SDSA model, after synthesis of DNA in the elongated chain, it moves to its original position and serves as a template to restore the lost sequence. As a result, only non cross-overs are originated (Paques and Haber, 1999; Symington, 2002).

*4.5.1.2 Break-Induced Recombination (BIR)*

Gene conversion usually occurs in short regions of DNA, gene conversion events involving wider areas of DNA can occur. In the BIR model (Fig. 13), after the invasion of the 3'-protruding chain, there is a stabilization of the displaced chain by alignment with the other end of the break. Consequently, the distal

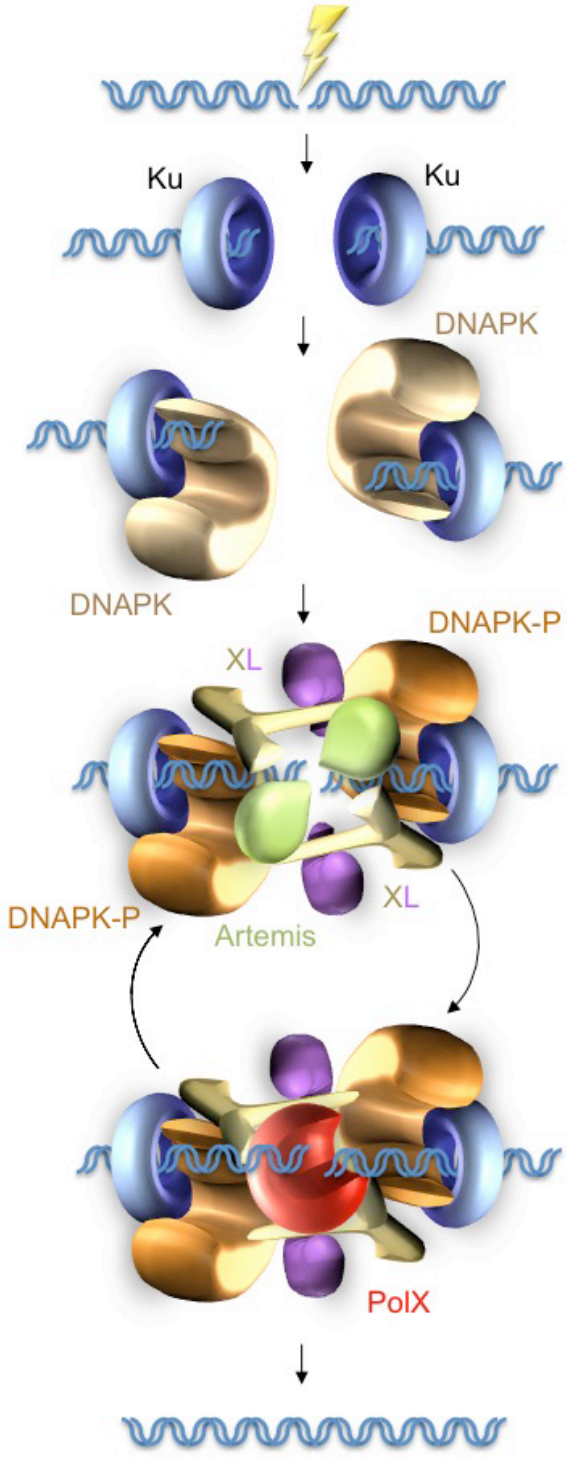



**Figure 14. Non-homologous end-joining pathway in eukaryotes.** This pathway acts repairing damage-generated DSBs. The Ku70/80 heterodimer is the first protein factor to arrived at the site of he break and bind the DNA ends. The DNA PKcs is the recruited and forms a complex with Artemis. The phosphorylated Artemis acts as an endonuclease, generating ssDNA protruding regions at the ends, and after this the complex dissociates from the DNA. The X family DNA polymerases are then in charge of searching for microhomologies or generating them, as well as filling in the gaps generated. Finallly, the XRCC4/Ligase IV complex seals the break.

break sequences are lost, replaced by the sequence of the donor chain. This process of non-reciprocal exchange of genetic material is known to be important in maintaining telomeres in the absence of telomerase (Paques and Haber, 1999; Symington, 2002).

*4.5.1.3 Single-Strand Annealing (SSA)*

This mechanism occurs when the DSB occurs between direct repeats (Fig. 13). After the break, if the single stranded formed is extensive enough to reach the complementary regions flanking the DSB, they can be aligned. The remaining single-stranded tails are removed by the action of endonucleases Rad1/Rad10 and gaps are filled by DNA polymerases and then are linked. This process results in the deletion of regions of repeated DNA region between the two repetitions, and thus is considered a mutagenic mechanism (Paques and Haber, 1999; Symington, 2002).

***4.5.2 Non-Homologous End-Joining (NHEJ)***

As mentioned, there are two types DSB repair: homologous recombination (HR) and non-homologous end-joining (NHEJ). The HR system requires that the cell is diploid level in the region to repair, because it uses the allele is intact in order to align and assemble the damaged ends by complementary sequences (West, 2003). Thus, the main advantage of this mechanism is its fidelity, as it guarantees an optimal repair, contributing to the maintenance of genomic stability. NHEJ system, however, is based on little or no homology between sequences to perform repair, since the proteins involved in the process recognize the ends of DNA based on their structure rather than its sequence (reviewed in {Lieber et al., 2004). Repair by this mechanism may lead to mutagenesis, contributing to the variability of the genomes (Ferguson et al., 2000; Heidenreich et al., 2003), and may be key to certain cellular processes such as antibody repertoire generation. NHEJ is the main mechanism to repair DSBs in higher eukaryotes, as it is operative throughout the cell cycle, unlike HR, which is inhibited during the phases G0, G1 and S in these organisms (Lieber et al., 2004).

The first step of this mechanism is the binding of specific protein factors to the ends of the DNA break (Fig. 14). The Ku70/Ku80 heterodimer recognizes the ends of the break, and due to its toroidal shape accommodates the duplex DNA, preventing possible nucleolytic degradation (Walker et al., 2001). Then, the DNA-PK kinase is recruited (Dvir et al., 1992; Gottlieb and Jackson, 1993), inducing a slight internalization of the Ku heterodimer (Dynan and Yoo, 1998), and allowing both sides of the break to approach through specific protein-protein interactions (Chen et al., 2000; DeFazio et al., 2002; Yaneva et al., 1997). Once the ends are juxtaposed, generally cannot be directly linked, but require pre-processing. Analysis of the sequences repaired by NHEJ at the break points suggests that some of these events involve the alignment of the ends through microhomologies (complementary sequences from 1 to 4 nt) near the site of rupture (Kramer et al., 1994; Moore and Haber, 1996; Roth and Wilson, 1986). When there is no direct microhomology the system must generate it by certain mechanisms that involve nucleases and/or DNA polymerases (Ma et al., 2002; Wilson and Lieber, 1999), which would be needed to process distortions, flaps or gaps that may arise as a result of the alignment of the chains (reviewed in {Hefferin and Tomkinson, 2005). The Ku-DNAPK complex recruits the proteins needed for processing and subsequent ligation of the ends. Artemis, an ssDNA 3'➔5' exonuclease, is activated through phosphorylation by DNA-PK (Ma et al., 2002). Polynucleotide kinase (PNK), which has kinase and phosphatase activities (Karimi-Busheri et al., 1999), may also intervene in end-processing (Chappell et al., 2002). If the ends at this point were compatible, the last step of the mechanism would be the recruitment of the XRCC4/LigaseIV complex by Ku, which would carry out the ligation of the ends (Grawunder et al., 1997; Schar et al., 1997; Teo and Jackson, 1997; Wilson et al., 1997). If, on the contrary the ends were not compatible, a DNA polymerase would then be needed, since its activity would be critical for filling the gaps generated during the alignment of the chains of DNA (Ramsden, 2011; Wilson and Lieber, 1999). The X family DNA polymerases Pol$\mu$ (Ma et al., 2002; Mahajan et al., 2002) and Pol$\lambda$ (Lee et al., 2004; Ma et al., 2004; Ramsden, 2011) have been involved in the NHEJ process. Pol$\mu$ could perform a template-independent polymerization to create the necessary complementary sequences (Juarez et al., 2006; Ma et al., 2002). Finally, after processing the ends, the complex formed by LigaseIV/XRCC4/ADN would be

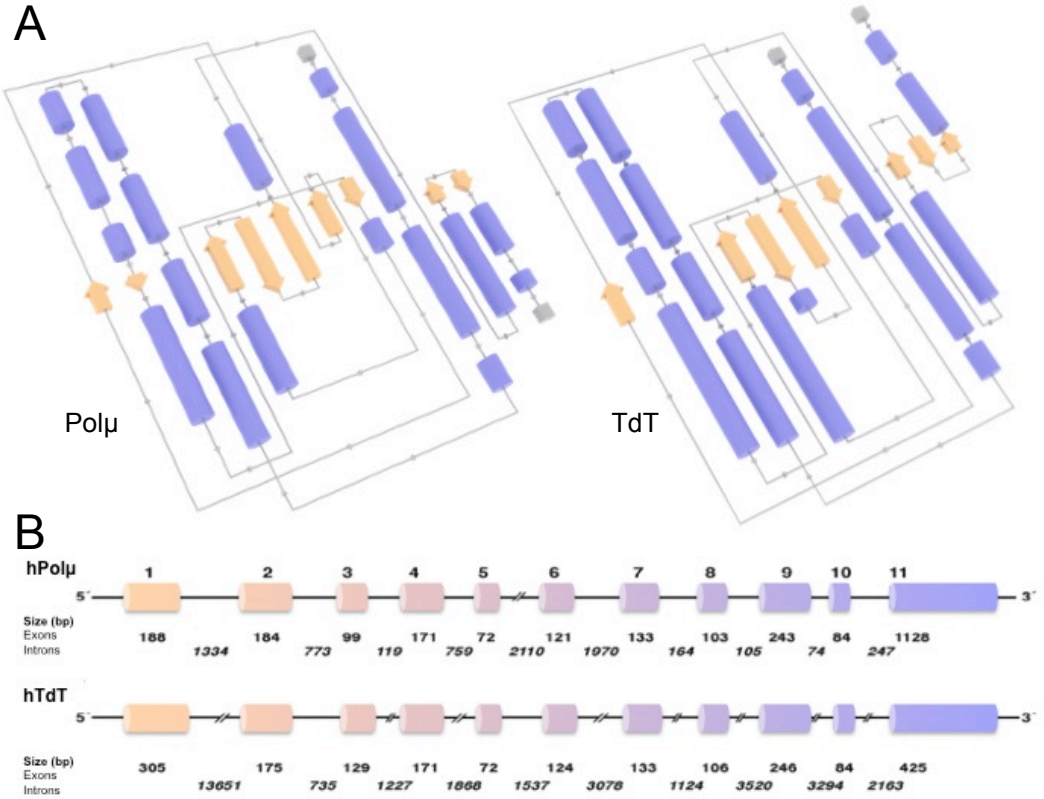



**Figure 15. Polμ and TdT similarities.**
A) Topological representation of the secondary structure elements of Polμ and TdT α-helices and ß-strands are represented as blue cylinders and orange arrows, respectively. B) Chromosomic organization of the human Polμ and TdT. Both DNA polymerases are coded by 11 exons (cylinders) separated by 10 introns (lines). The length of the exons and introns is indicated (bp).

responsible for sealing the joint between the ends of the break (Chen et al., 2000; Grawunder et al., 1997). Another factor similar to the protein XRCC4 has been recently identified in mammals. It has been called XLF (XRCC4-like factor) or Cernunnos, and interacts with LigaseIV/XRCC4/DNA complex to promote end ligation (Ahnesorg et al., 2006; Hentges et al., 2006).

*4.5.2.1 Alternative NHEJ*

A loosely defined alternative end-joining pathway operates in the absence of classical NHEJ factors such as Ku, XRCC4 or DNA ligase IV. These repair events frequently involve small deletions and entail short stretches of homology at the break point (Guirouilh-Barbat et al., 2004; Guirouilh-Barbat et al., 2007; Ma et al., 2003; Yan et al., 2007). This MMEJ (microhomology-mediated end-joining) pathway dominates during alternative end-joining. However, MMEJ and alternative end-joining are not synonymous, since error-free ligation can occur at low frequency in the absence of XRCC4-LigaseIV (Guirouilh-Barbat et al., 2007). Furthermore, MMEJ accounts for a proportion of V(D)J recombination events in wild-type cells (Corneo et al., 2007; Taccioli et al., 1993). MMEJ also appears to contribute to genomic instability in cancer. Translocation breakpoints in human cancers very often reveal microhomology, suggesting a role for MMEJ in translocation (Greenman et al., 2007). MMEJ may also facilitate chemotherapy resistance by genetic reversion in cells lacking wild-type *BRCA2* (breast-cancer susceptibility gene 2) (Edwards et al., 2008). In these cases, in-frame microhomologous deletions flanking the original mutation occurred in the resistant cells. The genetic requirements for MMEJ in cancer remain unclear (reviewed in {Hartlerode and Scully, 2009).

*4.5.2.2 Implicated Polymerases: Polμ*

Polμ is a DNA polymerase belonging to the X family that has been recently identified (Dominguez et al., 2000). This enzyme of 494 amino acids, and molecular weight of 55 kDa, has a strong to TdT, its closest counterpart in the X family (Fig. 15). They not only share 42% identity at the amino acid sequence, but also present a very similar structural organization: their N-terminal portion contains a nuclear localization sequence, followed by a BRCT domain and then the Polß-core structure already mentioned (Fig. 8). The BRCT domain was firstly identified in the BRCA1 protein, and has been shown involved in protein-protein interactions or protein-DNA in various repair and cell-cycle control proteins (Bork et al., 1997; Callebaut and Mornon, 1997).

Regarding Pol*μ* biochemical properties, it displays a certain terminal transferase activity (Dominguez et al., 2000), although it is primarily a DNA-dependent DNA polymerase (Dominguez et al., 2000; Ruiz et al., 2001) and its activity increases strongly in the presence of a template strand of DNA. It is also known that both types of polymerization is stimulated in vitro in the presence of $Mn^{2+}$ ions, the preferred metal activator, and in the presence of this cofactor Pol*μ* exhibits a strong mutator phenotype, with a very high probability of erroneous nucleotide incorporation, being one of the most error-prone polymerases known in higher eukaryotes (Dominguez et al., 2000). This strong mutator ability is based on a dislocation mechanism (Ruiz et al., 2004; Zhang et al., 2001) through which is capable of repositioning the template strand so that incorporation is dictated by templating bases away from the end of the primer. The mutator capacity of Pol*μ* is further enhanced by its low sugar discrimination, being able to incorporate not only dNTPs but also rNTPs (Nick McElhinny and Ramsden, 2003; Ruiz et al., 2003). This may have implications in cell cycle phases in which the levels of dNTPs are very low (Nick McElhinny and Ramsden, 2003) as rNTPs reserves remain high throughout the cycle.

As for its expression pattern, Pol*μ* is preferentially expressed in the germinal centers of secondary lymphoid organs (Dominguez et al., 2000), although Pol*μ* is also expressed at a basal level in every tissue. Although the in vivo role of Pol*μ* has not been clarified yet, a number of functions for the polymerase have been proposed, including its participation in the process of NHEJ.

*4.5.2.3 Implicated Polymerases: Polλ*

The newly identified Polλ is the X family member most similar to Pol (Aoufouchi et al., 2000; Garcia-Diaz et al., 2000; Nagasawa et al., 2000), since they share 32% identity at the amino acid sequence. Both polymerases have a similar organization of domains and three-dimensional structures (Fig. 8; {Garcia-Diaz et al., 2004), although Polλ presents an additional BRCT domain that is dispensable for catalysis (Nagasawa et al., 2000).

As Polß, Polλ shows a low processivity that increases in short DNA gaps with a 5'-phosphate, lacks 3'➔5' exonuclease activity, and contains an intrinsic dRP lyase activity, allowing it to replace Polß in short-patch BER reactions in vitro (Garcia-Diaz et al., 2001; Garcia-Diaz et al., 2002). Its high expression in testis together with its biochemical characteristics and the presence of the BRCT domain have led to propose a possible involvement of this polymerase in BER mechanisms during spermatogenesis (reviewed in Goodman, 2002). Recently this involvement has been demonstrated in vivo (Braithwaite et al., 2005). It has also been proposed a role for Polλ in long-patch BER, based on its limited ability of strand displacement in gapped-substrates (Garcia-Diaz et al., 2001). Recently, a possible role for Polλ in repairing DSBs has been proposed, specifically in the NHEJ pathway (Lee et al., 2004; Ma et al., 2004; Ramsden, 2011).

*4.5.2.4 Specialized functions of NHEJ*

The vertebrate immune system is characterized by intrinsic DSB production and repair as a mechanism of diversification of the B- and T-lymphocyte repertoire (reviewed in Dudley et al., 2005). Core members of the NHEJ pathway perform direct roles in V(D)J recombination. For example, Ku-deficient cells (Gu et al., 1997; Zhu et al., 1996) and DNA ligase IV-deficient cells (Frank et al., 1998) are defective in both coding and signal joint formation. Cells harbouring mutations in DNAPK are severely impaired in their ability to form coding joints, but show little or no defect in signal joint formation (Blackwell et al., 1989; Gao et al., 1998; Lieber et al., 1988). Artemis is also implicated in the formation of coding joints, but not signal joints (Nicolas et al., 1998; Rooney et al., 2002).

In contrast with V(D)J recombination, multiple DNA repair pathways are likely to be involved in CSR (class switch recombination), including base excision repair, mismatch repair and NHEJ (Chaudhuri and Alt, 2004). DNA sequences located between S (switch) regions can be detected in the form of excised circularized DNA, consistent with the involvement of DSB intermediates in CSR (Catalan et al., 2003; Iwasato et al., 1990). Sequences from CSR junctions show little or no sequence homology between donor and acceptor S regions, and often contain short deletions or insertions of nucleotides, all of which are consistent with DSB repair by NHEJ (Dunnick et al., 1993). Further evidence from knockout mice also suggests a role for NHEJ in CSR. DNAPK-deficient mice have significantly

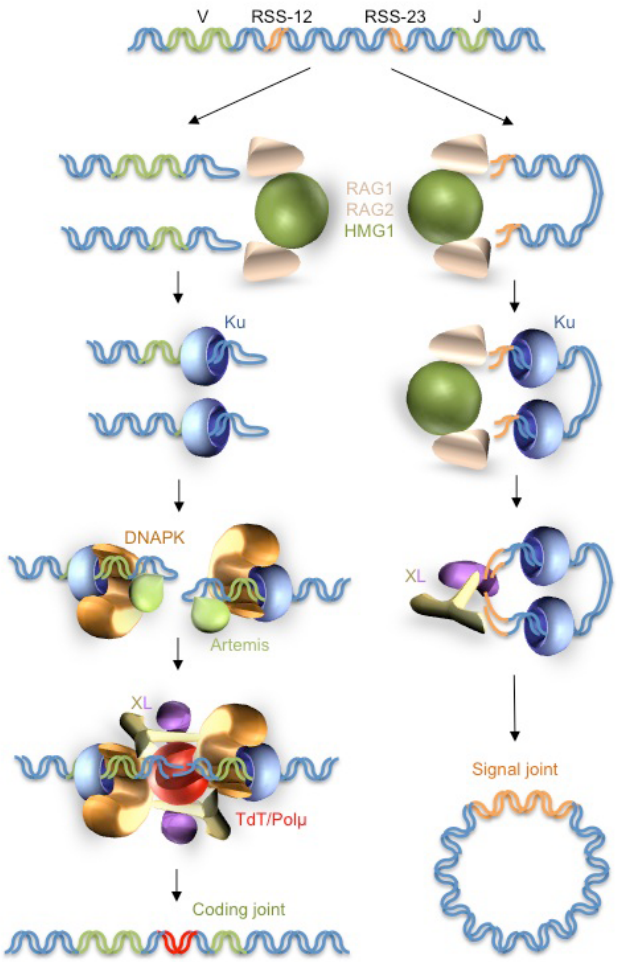



**Figure 16. V(D)J recombination.**
The recombination signal sequences (RSS), of 12 and 23 nts, are recognized by the Rag1/Rag2/HMG1 complex, which generates a break at these points. This provokes the recruitment of the NHEJ machinery, that acts bridging and repairing the DNA ends containing the V and J gene segments, as well as the ends of the fragment conatinin the RSS. Thus, two initially distant V and J segments end up being adjacently located.

reduced levels of serum Ig isotypes, and the only detectable isotype in Ku-deficient mice containing rearranged *IgH* and *IgL* genes is IgM (Manis et al., 2002; Manis et al., 1998).

NHEJ also plays a role in telomere biology (reviewed in {Riha et al., 2006). The formation of dicentric chromosomes as a consequence of DNA end-joining is a hallmark of telomere dysfunction. NHEJ appears to play a central role in the formation of dicentric chromosomes in cells with telomere dysfunction, since fusion of uncapped telomeres is strictly DNA ligase IV-dependent (Smogorzewska et al., 2002). In addition, Ku, DNAPK and the MRN complex participate in multiple facets of normal telomere biology. All three components of the MRN complex bind telomeres, and disruption of the MRN complex leads to telomere length deregulation (Boulton and Jackson, 1998; Ranganathan et al., 2001). The Ku heterodimer and DNA-PK also play roles in the regulation of normal telomere length (Boulton and Jackson, 1998; d'Adda di Fagagna et al., 2001; Espejel et al., 2004; Hande et al., 1999; Myung et al., 2004).

*4.5.2.5 V(D)J Recombination*

This mechanism, which is the first step during generation of antibody diversity, occurs in the precursors of T and B cells in the thymus and bone marrow, respectively. Involves recombination of gene segments encoding the variable region of antigen receptors (Tonegawa, 1983).

Immunoglobulins and T receptor are composed of two heavy and two light chains. The exon encoding the variable region of these originate from the fusion of gene segments V, D and J in the case of heavy chains and the fusion of V and J segments in the case of light chains. The diversity at this level is generated by combining different gene segments and by adding N nucleotides (not templated) in the binding sites between them (Tippin et al., 2004). As shown in figure 16, V(D)J recombination begins with the action of two lymphoid tissue endonucleases, Rag1 and Rag2 (Mombaerts et al., 1992; Oettinger et al., 1990; Schatz et al., 1989; Shinkai et al., 1992), which form a complex with a group of high mobility proteins (HMG1) (Sawchuk et al., 1997; van Gent et al., 1997), which introduces DSBs in recombination-specific sequences (RSS) flanking the multicopy segments V, D and J (Lieber et al., 2003). As a result of the cut two phosphorylated blunt ends are generated and form hairpins, while the fragment containing the gene segment forms another covalently closed hairpin (Lieber et al., 2004). Then the NHEJ machinery comes in to repair DSBs (reviewed in Lieber et al., 2003): the Ku protein complex binds to the end of each coding fragment and recruits DNA-PKcs. The DNA-PK complex thus formed can activate Artemis through phosphorylation, so that it is then able to open the hairpins in the coding ends, generating ssDNA segments (Ma et al., 2002). These segments can be used by TdT, which is recruited by Ku proteins (Mahajan et al., 1999), and adds untemplated nucleotides generating variability in the junction area (Desiderio et al., 1984). Finally, the LigaseIV/XRCC4

seals the joints (Chen et al., 2000; Grawunder et al., 1997; Lieber et al., 2004).

*4.5.2.6 TdT in V(D)J Recombination*

TdT is a very peculiar member of the X family. Identified in 1960, was the first known DNA polymerase capable of carrying out elongation of a single stranded 3'-OH in the absence of a template strand (Bollum, 1960; Kato et al., 1967).

The alignment of its amino acid sequence with that of Polß indicates that they share 22% identity at the catalytic domain, although TdT has a BRCT domain implicated in protein-protein interactions (Bork et al., 1997; Callebaut and Mornon, 1997) and an additional N-terminal region of 13 kDa, which contains a nuclear localization sequence (Fig. 8).

The recent solving of the three-dimensional structure of TdT complexed with DNA (Delarue et al., 2002) has led to propose the structural basis of their preference for DNA substrates without template strand. Among its biochemical characteristics it is worth noting its ability to incorporate both rNTPs and dNTPs in vitro (Boule et al., 2001; Roychoudhury, 1972) and a long list of unnatural nucleoside triphosphates (Krayevsky et al., 2000; Semizarov et al., 1997). TdT expression, strongly restricted to primary lymphoid organs (thymus and bone marrow), and its ability to polymerize in the absence of a template strand, allowed to engage this polymerase in the process of generating variability during V(D)J recombination, specifically at the level of the incorporation of untemplated nucleotides (N additions) between segments V, D and J, as shown by the analysis of these regions in mice deficient in this polymerase (Komori et al., 1993).

*4.5.2.7 Polμ in V(D)J Recombination*

Polμ preferential expression in lymphoid tissues, especially in the germinal centers of secondary lymphoid organs, suggesting a specific role of this polymerase in processes occurring in these regions. Its resemblance to TdT at the structural level, and its ability to conduct untemplated nucleotide additions, together with the fact that TdT is not expressed in secondary lymphoid organs, allowed to propose a function for Polμ in V(D)J recombination in the germinal centers (Ruiz et al., 2001), which occurs in these regions as an additional mechanism for inactivation or diversification of the immune response (Papavasiliou et al., 1997). Moreover, Polμ

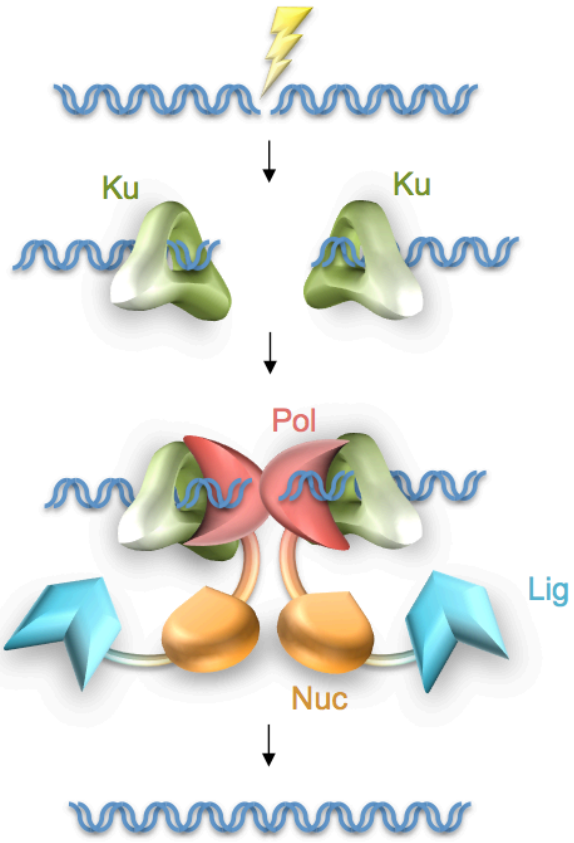



**Figure 17. The NHEJ repair pathway in prokaryotes.**
After the binding of Ku to 3'-protruding ends and further recruitment of LigD, it specifically recognizes the 5' phosphate and directly mediates the synapsis event required for end-joining. After microhomology pairing, 3'-protruding termini can be resected by the NucDom of LigD. Once the required nucleolysis has occurred, resynthesis by PolDom and ligation of nicks complete break repair.

is present also in the thymus and bone marrow, and thus may be required during the normal process of V (D)J recombination as DNA-dependent polymerase to generate palindromic sequences (P sequences) at the ends of the coding fragments, or during gap-filling reactions required for coupling N additions to the DNA ends (Ruiz et al., 2001). It was recently demonstrated in vivo role of Polμ the VJ recombination process of the light chain (kappa) of immunoglobulins, based on the observed deletions at the junctions between these gene segments in the case of Polμ deficiency (Bertocci et al., 2003). Also, recent data implicated Polμ in the V(D)J recombination in mice embryos, a stage in which TdT is still not expressed (Gozalbo-Lopez et al., 2009). In this case, all the N-additions observed in wild type mice were completely attributable to Polμ, as shown by comparison with Polμ KO mice. This evidence suggests a role for Polμ in the V(D)J mechanism.

### *4.5.3 DSB repair in prokaryotes*

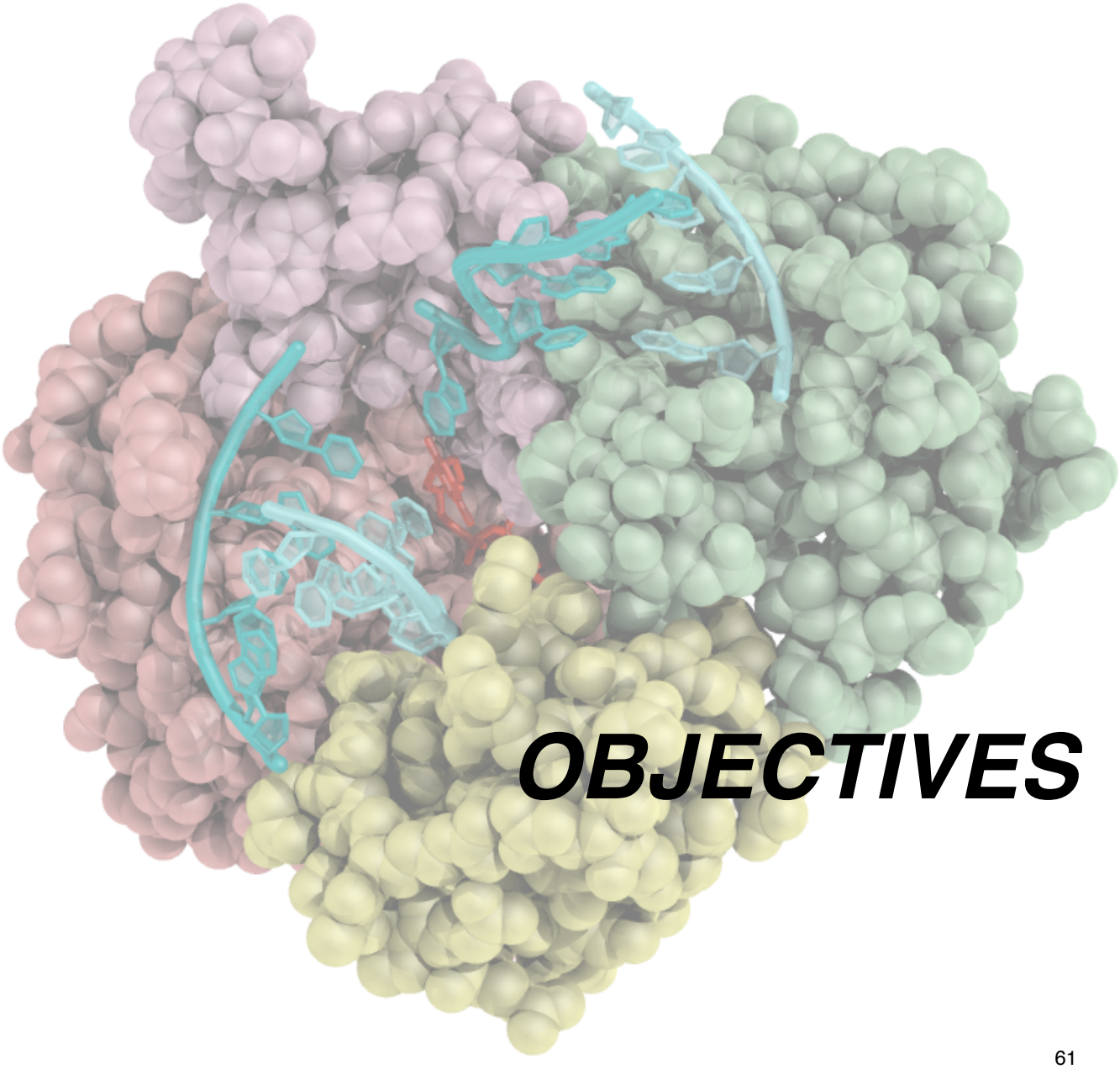
NHEJ was first identified in mammalian cells and assumed to be restricted to the eukarya. The eukaryotic NHEJ pathway is defined by its essential constituents: the Ku70/80 heterodimer, DNA-PKcs (absent in yeast) and the ligase IV/XRCC4/XLF (LXX) complex (Daley et al., 2005; Mahaney et al., 2009). A host of other proteins have also been implicated in eukaryotic NHEJ, including Pol*μ* and Polλ (both X family DNA polymerases).

More recently, a functionally homologous NHEJ repair apparatus has been identified and characterized in prokaryotes (Della et al., 2004; Pitcher et al., 2007; Weller et al., 2002). The mycobacterial NHEJ complex is composed of a Ku homodimer and a multifunctional DNA ligase D (LigD), which together are capable of recognizing, processing and resealing DSBs (Fig. 17; Della et al., 2004). This primordial NHEJ complex is present in the majority of bacterial species (Bowater and Doherty, 2006), suggesting that this end-joining pathway is physiologically important for genome maintenance in prokaryotes. Indeed, NHEJ-deficient stationary phase cells (Mycobacteria) and spores (Bacillus) are sensitive to agents that produce DSBs (i.e. ionizing radiation, dry heat and desiccation) (Gong et al., 2005; Pitcher et al., 2007; Weller et al., 2002), indicating that NHEJ is also predominantly utilized by non-replicating cells for break repair.

In addition to its core DNA ligase domain, the mycobacterial LigD also comprises auxiliary polymerase (PolDom) and nuclease domains (NucDom), raising the possibility that LigD encodes a self-sufficient NHEJ repair machine (Aravind and Koonin, 2001; Weller and Doherty, 2001; Weller et al., 2002). Ectopic co-expression of the *Mycobacterium tuberculosis* (*Mt*) Ku and LigD genes complements NHEJ-defective yeast strains and restores DSB repair (Della et al., 2004). Furthermore, bacterial and yeast NHEJ resulted in different molecular signatures at processed joints, indicating that bacterial NHEJ processing steps are distinct, thereby indirectly implicating the PolDom and NucDom activities. These results suggest that Ku and LigD possess all of the break-recognition, end-processing and ligation activities required to facilitate the complex task of DSB repair. An array of *in vitro* biochemical studies also support the proposal that Ku and LigD form a two-component DNA repair complex at the termini of DSBs (Della et al., 2004; Weller et al., 2002), in addition to the observation that Ku and LigD are genetically associated in many bacterial species (Aravind and Koonin, 2001; Della et al., 2004; Weller and Doherty, 2001; Weller et al., 2002).

The prokaryotic NHEJ polymerases (PolDom) are members of the archaeo-eukaryotic primase (AEP) superfamily (Aravind and Koonin, 2001; Weller and Doherty, 2001). Historically, primases have been regarded as specialized polymerases with a role restricted to the synthesis of RNA primers at replication forks, which act to “prime” DNA extension by replicative polymerases (Frick and Richardson, 2001). However, this limited role for AEPs in DNA metabolism has been challenged by the discovery of a wide variety of structurally homologous bacterial AEPs (Pitcher et al., 2007; Zhu et al., 2006). These enzymes possess multiple nucleotidyl transferase activities, including DNA-dependent RNA primase, terminal transferase and DNA-dependent DNA/RNA gap-filling polymerase activities (Della et al., 2004; Pitcher et al., 2007; Zhu and Shuman, 2005). In addition to these activities, a reduced template-dependence and flexible active site enables these nucleotidyltransferases to generate template distortions, primer realignment and lesion bypass, which greatly enhances their role in DSB repair processes (Pitcher et al., 2007; Yakovleva and Shuman, 2006). The termini of DSBs are often damaged *in vivo*, consequently end-alignment, nucleolysis and gap-filling reactions are required to remodel mispaired or damaged nucleotides.

Uniquely among the three domains that constitute the mycobacterial LigD, the polymerase domain physically interacts with Ku (Pitcher et al., 2005), suggesting that LigD recruitment to DSBs *in vivo* is accomplished primarily *via* this interaction. Ku appears to also enhance polymerase extension activity. The prokaryotic NHEJ polymerases can directly promote the connection (synapsis) of 3' overhanging ends of a DSB (Brissett et al., 2007). Thus, when long DNA 3’-protrusions are not fully complementary, a PolDom-mediated stable but imperfect synapsis is proposed to be a necessary bridging step allowing end-resection prior to gap filling and ligation.



# OBJECTIVES

During the present PhD Thesis our fields of interest have been the following:

1. Searching for the DNA binding properties of the family X polymerases with a function in NHEJ: by EMSA assays and optimized footprinting assays we set out to study the importance of the 5'-P group of the downstream strand and the ability of the different PolXs to bind gapped and 3'-protruding substrates.

2. Handling of two DNA substrates by Pol*μ* during NHEJ: evaluating the impact of the sequence context, the presence of gaps or nicks, and the size and separation of these gaps in the connection of two DNA ends, and also the ability of Pol*μ* to deal with non-aligned ends containing flipped-out bases, mismatches, flaps, etc., and oxidative damages like the modified base 8oxoG.

3. Importance of the BRCT domain and the "brooch" motif (YACQR) during Pol*μ*-mediated NHEJ: by using site-directed mutagenesis (point and deletion mutants) we focused on testing the effect of these structural features on the DNA binding and polymerization activities of Pol*μ*.

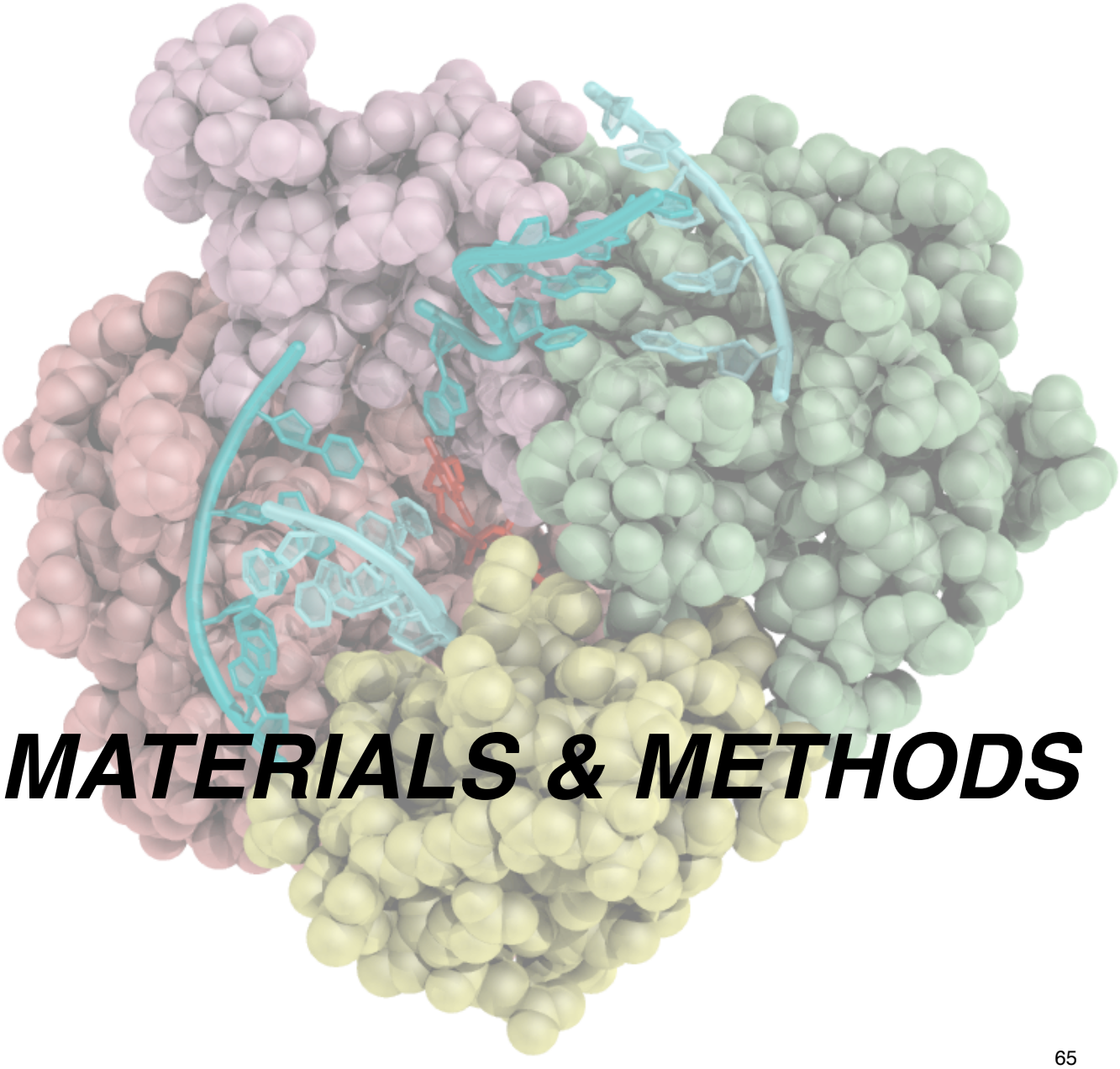
4. Substrate usage by human Pol*μ* during gap-filling and NHEJ reactions: by evaluating the consequences of the use of ribonucleotides as opposed to deoxynucleotides as substrates, and of different activating metal ions, specifically $Mn^{2+}$, both *in vitro* and in human and yeast cell extracts.

5. Analysis of Pol*μ* residues involved in NHEJ: by site-directed mutagenesis we wanted to find the residues or regions of Pol*μ* that specifically increase the efficiency of the end-joining reaction, and to this end we decided to mutate residues implicated in the binding to the primer or template DNA strands.

6. Structural and functional relationship between the terminal transferase and the end-joining activities of Pol*μ*: we set out to determine if the exceptional ability of Pol*μ* to insert untemplated nucleotides was of use during the NHEJ reactions, and to perform an in depth study of the intimate mechanism of these two activities. For this we used deletion and single mutants to examine the roles of two loops present in Pol*μ*: Loop 1 and the previously uncharacterized thumb "mini-loop".

7. During the course of this PhD Thesis a number of evidences concerning the post-translational regulation of several family X members came to light. Thus, a later objective was to determine if Pol*μ* was under similar control mechanisms, by using both biochemical (by means of *in vitro* assays) and cellular approaches.

8. As an ongoing collaboration with Prof. Doherty, we continued the study of the enzymatic properties of the polymerase domain (PolDom) of the multifunctional protein LigD from *Mycobacterium tuberculosis*, implicated in the bacterial NHEJ pathway. Site-directed mutagenesis approaches, combined with specific gap-filling and NHEJ *in vitro* assays, were selected to analyze two different complexes obtained by X-ray crystallography: a catalytically competent pre-ternary complex, which lacks the primer strand, and a fully complementary synapsis of two DNA ends mediated by a PolDom dimer.



# MATERIALS & METHODS

**1. Reagents**

Ultrapure NTPs, dNTPs, ddNTPs, [α-$^{32}$P]dNTPs (3000 Ci/mmol) and [γ-$^{32}$P]ATP (3000 Ci/mmol) were purchased from GE Healthcare (USA). T4 polynucleotide kinase, purified TdT and the CDK/cyclin complexes were obtained from New England Biolabs (Beverly, MA, USA). Pfu DNA polymerase and DNAse I were purchased from Promega Corporation (Madison, WI, USA). Purified *Mt*-PolDom wild type and mutant versions were kindly provided by Prof. A.J. Doherty. Mutant Pol*μ*-ΔBRCT was obtained as described in Juarez, 2006. Mutant Pol*μ*-Δloop was obtained as described in Juarez et al., 2006. Polλ was obtained as described in Garcia-Diaz et al., 2000. Pol*μ* mutants G433Y, W434F and W434R were obtained as described in Ruiz et al., 2003.

**2. DNA Oligonucleotides**

Synthetic DNA oligonucleotides were obtained from Isogen, Invitrogen, Sigma and Eurogentec. In each case, the indicated PAGE or HPLC-purified oligonucleotides were labeled at their 5' ends with [γ-$^{32}$P]ATP for detection during biochemical assays. The oligonucleotides used to generate the DNA substrates are listed in Table 1.

**3. Construction and Purification of Human Polß**

Polß cDNA was obtained from the IMAGE consortium and subcloned in the pET16b bacterial overexpression plasmid. The final clone was transformed in *E. coli* BL21(DE3)pRil, cells were grown up to 0.6 D.O. and Polß expression was induced with 0.5 mM IPTG during 160 min. Cells were lysed with alumina and centrifuged for 15 min at 15.000 rpm to obtain a debris- and insoluble protein-free extract. DNA was precipitated with poliethylenimine and eliminated by centrifugation. Soluble proteins in the supernatant were precipitated with 65% ammonium sulfate. The pellet was resuspendend and subjected to chromatography in a 5 ml phosphocellulose column, followed by chromatography in a 5 ml heparin HiTrap column (Pharmacia Biotech). After a final concentration step in a 1 ml phosphocellulose column, the fractions containing highly purified Polß were stored at -80ºC.

**4. Construction and Purification of Human Pol*μ* and mutant versions.**

Site-directed mutagenesis by a single PCR with oligonucleotides containing the desired mutation was performed on the overexpression plasmid for Pol*μ* (pRSETA-hPol*μ*; Ruiz et al., 2001). DNA constructs were sequenced and transformed in *E. coli* BL21(DE3)pLysS. Wild-type and mutant Pol*μ* variants were over-expressed and purified using an Äkta Purifier FPLC system (GE Heatlhcare). Briefly, the bacterial pellet was sonicated, centrifuged to eliminate cell debris, filtered and loaded on a heparin column. The eluted fractions containing the protein of interest were loaded on an SP Sepharose column, and the resulting Pol*μ*-containing fractions loaded on a size exclusion chromatography column. Pol*μ*-ΔBRCT mutant was purified using the Äkta Purifier FPLC system with a slightly different protocol: the cleared bacterial lysate was loaded on S and Q Sepharose columns in tandem, the flow-trough was then subjected to a second chromatography step on a heparin column, and the eluted fractions were loaded on a manually packed phosphocellulose column to eliminate contaminant nucleases. In both cases, the eluted fractions containing highly purified protein were concentrated and stored at -80ºC. The oligonucleotides used for the site-directed mutagenesis are listed in Table 2.

**5. Biochemical Assays**

**5.1 Electromobility shift assays**

EMSAs, used to analyze the interaction with gapped-substrates and NHEJ intermediates, were performed in a final volume of 12.5 *μ*l, containing 50 mM Tris–HCl (pH 7.5), 0.1 mg/ml of BSA, 1 mM DTT, 4% glycerol, 5 nM labelled DNA and different concentrations of the indicated proteins. After incubation for 10 min at 30ºC samples were mixed with 3 *μ*l of 30% glycerol and resolved by native gel electrophoresis on a 4% polyacrylamide gel (80:1 [w/w] acrylamide/bisacrylamide). Gels were dried and labelled DNA fragments were detected by autoradiography.

**5.2 Analysis of the interaction with single stranded DNA by nitrocellulose filter binding.**

The assay was carried out using 5 nM of a labeled homopolymer (PolydT) as DNA primer, containing a ddCTP in the 3'-end to avoid polymerization, in the presence of 1 mM $MnCl_2$ and the indicated concentration of either Pol$\mu$ wild type, R387K, R387A or H329G. When pointed out, 100 $\mu$M dTTP was added. Reactions were incubated for 10 min at 22ºC in binding buffer (50 mM TrisCl pH 8.0, 25 mM NaCl, 1 mM DTT), and filtered on nitrocellulose filters as described (Wong and Lohman, 1993). Radiolabeled DNA retained in each filter was quantitated by Phosphorimager.

**5.3 DNA Footprinting Assays.**

The indicated proteins at the designated concentrations were incubated with 30 nM labelled gapped substrate in a final volume of 20 $\mu$l, containing 50 mM Tris–HCl (pH 7.5), 1 mM DTT, 4% glycerol and 0.1 mg/ml of BSA. After incubation for 10 min at 37°C, samples were treated with 0.03 units of commercial DNAseI for 2 min at 37ºC. Reactions were stopped with a buffer containing 20 mM EDTA, and the DNA precipitated with 3 M sodium acetate and 100% EtOH, O/N at -80ºC. The DNA pellets were washed with 70% EtOH and resuspended in loading buffer (10 mM EDTA, 95% [v/v] formamide, 0.03% [w/v] bromophenol blue, 0.03% [w/v] xylene cyanol), boiled and subjected to electrophoresis in 8 M urea-containing 8% polyacrylamide sequencing gels. Labelled DNA fragments were detected by autoradiography.

**5.4 DNA Polymerization Assays.**

To analyze DNA-dependent and independent polymerization, several DNA substrates, containing 5'-P-labeled primers, were incubated with different proteins, at the concentration indicated in each case. The reaction mixture, in 20 $\mu$l, contained 50 mM Tris–HCl (pH 7.5), 1 mM DTT, 4% glycerol and 0.1 mg/ml BSA, in the presence of the indicated amounts of the DNA polymerization substrates, and of NTP/dNTP and activating metal ions. After incubation for 30 min at 30ºC, reactions were stopped by adding gel loading buffer (95% [v/v] formamide, 10 mM EDTA, 0,1% [w/v] xylene cyanol and 0.1% [w/v] bromophenol blue) and analyzed by 8 M urea-20% PAGE and autoradiography. When indicated, we used ddNTPs instead of dNTPs to limit incorporation to a single nucleotide on the 3'-end of the labeled oligonucleotide.

**5.5 NHEJ Assays.**

NHEJ polymerization assays were carried out essentially as described above for polymerization assays, but using independent DNA primer (labeled) and template (unlabeled) molecules. When indicated, we used variable concentrations of ddNTPs instead of dNTPs/NTPs to limit incorporation to a single nucleotide on the 3′-end of the labeled oligonucleotide.

**6. *In vitro* kinase assays and mass-spectrometry analysis.**

Phosphorylation was carried out in a final volume of 15 $\mu$l containing kinase buffer (50 mM Tris, pH 7.5, 10 mM MgCl2, 1 mM EGTA, 1 mM DTT and 100 $\mu$M ATP, 0.3 $\mu$Ci of [$\gamma$-$^{32}$P] ATP), in the presence of substrate (either Histone H1 or 700 ng of Pol$\mu$ WT or the different mutants) and 5 units of the indicated Cdk/cyclin complexes (New England Biolabs). Reactions were carried out for 60 min at 30ºC, and the proteins separated on a 10% SDS–PAGE. Reactions were stopped by addition of loading buffer and boiling for 5 min at 95ºC. The gel was stained by Coomassie Brilliant Blue, dried and exposed to an X-ray film. When kinases assays were performed for mass spectrometry, only cold ATP was used.
For the reverse-phase liquid chromatography mass-spectrometry (RP-LC-MS/MS) analysis, gel bands were digested in situ with trypsin (5 ng/$\mu$l; Promega) for 1 hour in an ice-bath as described in Shevchenko et al., 1996, with minor modifications. Stained protein gel bands were incubated in 50 mM $NH_4HCO_3$ (BA) prior to reduction with 10 mM DTT and alkylation with 55 mM iodoacetamide, both in 50 mM BA. Digestion buffer was removed and gels were covered again with 50 mM BA and incubated for 12 hours at 37ºC. Whole supernatants were allowed to dry and then stored at -20ºC until mass spectrometry analysis was performed. The tryptic peptides were resuspended in 9 $\mu$l of 0.1% formic acid and analyzed by RP-LC-MS/MS in a Surveyor HPLC system coupled to an Ion Trap Deca XP mass spectrometer (Thermo Fischer Scientific). The peptides were separated by reverse phase chromatography using a 0.18 mm x 150 mm Bio-

Basic C18 RP column (Thermo Fischer Scientific), operating at 1.8 µl/min. Peptides were eluted using a 50 min gradient from 5 to 40% solvent B (Solvent A: 0.1% formic acid in water; Solvent B: 0.1% formic acid, 80% acetonitrile in water). ESI ionization was done using a microspray "metal needle kit" (Thermo Fischer Scientific) interface. The mass spectrometer was operated in the MS/MS ion monitoring mode. In this mode, Deca XP detector was programmed to perform, along the entire gradient, a continuous sequential operation in the MS/MS mode on the doubly charged ions corresponding to the peptide/s selected, using an isolation width of 2 µ (in mass to charge radio units), normalized collision energy of 35%. The MS/MS spectra from the peptide was analyzed by assigning the fragments to the candidate sequence, after calculating the series of theoretical fragmentations, according to the nomenclature of the series as previously described (Roepstorff and Fohlman, 1984).

**7. Amino acid sequence comparisons and 3D-modelling.**

Multiple alignment of different DNA polymerases was done using the program MULTALIN (http://prodes.toulouse.inra.fr/multalin/). The different conformations of the studied residues, motifs and domains in the X family and bacterial polymerases were analyzed with the software MacPymol (http://delsci.com/macpymol/). 3D modelling of polymerases of unknown atomic structure was performed firstly using the protein structure prediction webserver Phyre2 (Kelley and Sternberg, 2009) to select the best candidates from the Protein Data Bank (PDB) and then using the structure homology-modeling SwissModel webserver (Arnold et al., 2006; Kiefer et al., 2009; Peitsch et al., 1995).

**8. Cell culture and preparation of nuclear cell extracts.**

HeLa cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (GIBCO BRL) and penicillin/streptomycin (Sigma). To obtain nuclear cell extracts, 20 plates (150 mm) were harvested by scraping in PBS. Cells were washed with PBS and resuspended in 4 PVC (packed cell volume) hypotonic buffer (10 mM Cl, 10 mM Tris HCl pH 7.9, 1 mM DTT, 1 µg/ml pepstatin, 1 µg/ml leupeptin, 1 µg/ml soybean trypsin inhibitor, 10 µg/ml PMSF) for 10 min on ice. After incubation, cells were lysed with a Dounce homogenizer (20 strokes). Nuclei were pelleted at 2000 rpm for 10 min and washed in hypotonic buffer prior to resuspension in 4 PVC nuclear extraction buffer (50 mM Tris HCl pH 7.9, 0.42 M KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 20% glycerol, 10% sucrose, 2 mM DTT, 1 µg/ml pepstatin, 1 µg/ml leupeptin, 1 µg/ml soybean trypsin inhibitor, 10 µg/ml PMSF). After incubation with agitation for 30 min at 4ºC, the samples were centrifuged for 30 min at 15.000 rpm and the supernatants collected. To precipitate the soluble proteins, 0.33 g ammonium sulfate per ml and 10 µM of NaOH 1M (per gram of ammonium sulfate) were added. After incubation for 20 min at 4ºC and centrifugation for 10 min at 13.300 rpm, protein pellets were resuspended in 0.25 PVC of storage buffer (20 mM Tris HCl pH 7.9, 0.1 M KOAc, 0.5 mM EDTA, 20% glycerol, 1 mM DTT, 1 µg/ml pepstatin, 1 µg/ml leupeptin, 1 µg/ml soybean trypsin inhibitor, 10 µg/ml PMSF) and dialyzed against same buffer over/night at 4ºC. After a final centrifugation at 85.000 rpm for 1 hour, aliquots were snap-frozen in liquid nitrogen and stored at -80ºC.

**9. Yeast culture, protein overexpression and preparation of whole cell extracts.**

Strains of *Schizosaccharomyces pombe* were grown in YES rich medium (yeast extract 0.5%, 3% glucose, supplemented with 225 mg/ml adenine, histidine, leucine, uracil and lysine). Strains transformed with plasmids whose selection marker gene is an auxotrophy were selected on EMM (Edinburgh minimal medium). To repress the nmt1 promoter of pREP41 and pDS473a vectors, 15 mM thiamine was added to the medium. For growth on solid medium, 2% agar was added. The yeast strains were made competent by the lithium acetate method and transformed by heat shock as described in Moreno et al., 1991.

To obtain whole cell extracts, strains of *S. pombe* were grown in 50 ml of YES to reach log phase (0.4-0.8 D.O.). Subsequently, cells were pelleted at 4ºC and washed with stop buffer (150 mM NaCl, 50 mM NaF, 10 mM EDTA and 1 mM $NaN_3$, pH 8). The cell pellet was frozen at -80ºC. Thawed cells were resuspended on ice in 200 µl of cold lysis buffer

(PBS, 50 mM NaF, 2 mM EDTA pH 8, 1% NP-40, p-$NH_2$-benzamidine 1.3 mM, 1 mM PMSF and protease inhibitors from Roche) and lysed with the same volume of glass beads (425-600 microns, Sigma) at 4ºC with 3x 15 second pulses in a Fast-Prep (BioSavant). Lysate was collected, centrifuged at 15.000 rpm for 30 min at 4ºC to remove cell debris and the soluble fraction was collected and stored at -80ºC. The protein concentration was determined by Protein Assay (Bio-Rad).

**10. Western Blot.**

Cells were lysed, and extracted proteins were resolved in 13% SDS-polyacrylamide gels and transferred to nitrocellulose membranes. Endogenous Pol*μ*, Polλ, Ku70, XRCC4 and Ligase IV were detected by specific rabbit polyclonal (α-Pol*μ* and α-Polλ, own source) and mouse monoclonal antibodies (sc9033, sc5606 and sc11750, Santa Cruz Biotechnologies). Over-expressed Pol4, hPol*μ* and hPolλ were detected with anti-GST monoclonal antibody (G1160, Sigma). As a loading control, α-actin was detected using a rabbit polyclonal antibody (Sigma). Membranes were stripped with stripping buffer (recipe), for 30 min at 60ºC. Proteins were visualized by enhanced chemiluminescence detection (Amersham Biosciences) using goat anti-mouse or anti-rabbit IgGs coupled to horseradish peroxidase as the secondary antibody (Amersham Biosciences).

**11. Complex crystal preparation.**

*Mt*-PolDom was expressed and purified as described in Pitcher et al., 2005. The *Mt*-PolDom-DNA-UTP complex was prepared by taking the previously described crystals (Brissett et al., 2007) and soaking in drops of mother liquor containing UTP and $MnCl_2$ to final concentrations of 10 mM and 20 mM respectively. The crystals were harvested and cryo protected in reservoir buffer plus 17% ethylene glycol before snap freezing in liquid nitrogen. All data sets were collected at 100K. Single wavelength diffraction data of *Mt*-PolDom-DNA were collected at beam-line BM-16 of the ESRF (Grenoble, France). The crystals belong to space group $P4_1$ unit cell dimensions: $a$ = 145.697Å, $b$ = 145.697Å, $c$ = 44.953Å, α = β = γ = 90º. The diffraction data were processed with HKL2000 (Otwinowski et al., 2003) with additional processing by programs from the CCP4 suite (CCP4, 1994). Done in collaboration with Prof. Doherty laboratory (Univ. Sussex).

**12. Structure solution and refinement of *Mt*-PolDom pre-ternary complex.**

The structure of the PolDom/DNA/UTP complex was determined by molecular replacement using the program PHASER (McCoy et al., 2005). The crystallographic model of (apo) *Mt*-PolDom (PDB ID: 2IRU) was used as a molecular replacement search model. Initial refinement was carried out against 95% of the data with REFMAC5 (Murshudov et al., 2011). The remaining 5%, which were randomly excluded from the full data set, was used for cross-validation by calculating the $R_{free}$ to follow the progress of the refinement. The same subset of reflections was used throughout the refinement. A final refined model at 3.0Å resolution, with an $R_{cryst}$ of 17.96% and $R_{free}$ of 23.19%, was obtained. Each cycle of refinement was accompanied by manual rebuilding using the program COOT (Emsley and Cowtan, 2004). Done in collaboration with Prof. Doherty laboratory (Univ. Sussex).

**13. Steady-state fluorescence emission experiments.**

Emission fluorescence data for 2-aminopurine (2AP) DNAs and *Mt*-PolDom complexes were collected using a Varian Cary Eclipse fluorescence spectrophotometer. Samples were excited at 360 nm and fluorescence emission data scans were collected between 340-470nm. Bandpass slits were 10 and 20 nm for excitation and emission channels respectively. Solutions containing 200 nM 2AP-labelled DNA, 200 nM *Mt*-PolDom, 50 mM Tris (pH 7.5), 150 mM NaCl, 10 mM $MnCl_2$, and 4% glycerol were titrated with stock solutions of UTP. Intrinsic protein fluorescence was subtracted. Fluorescence values were plotted and relative quenching was fitted to the equation, $F = ( [UTP] / K_d [UTP] ) + Ns [UTP]$ where, F is the relative quenching specific to the complex, and *Ns* is the non-specific binding constant. Done in collaboration with Prof. Doherty laboratory (Univ. Sussex).

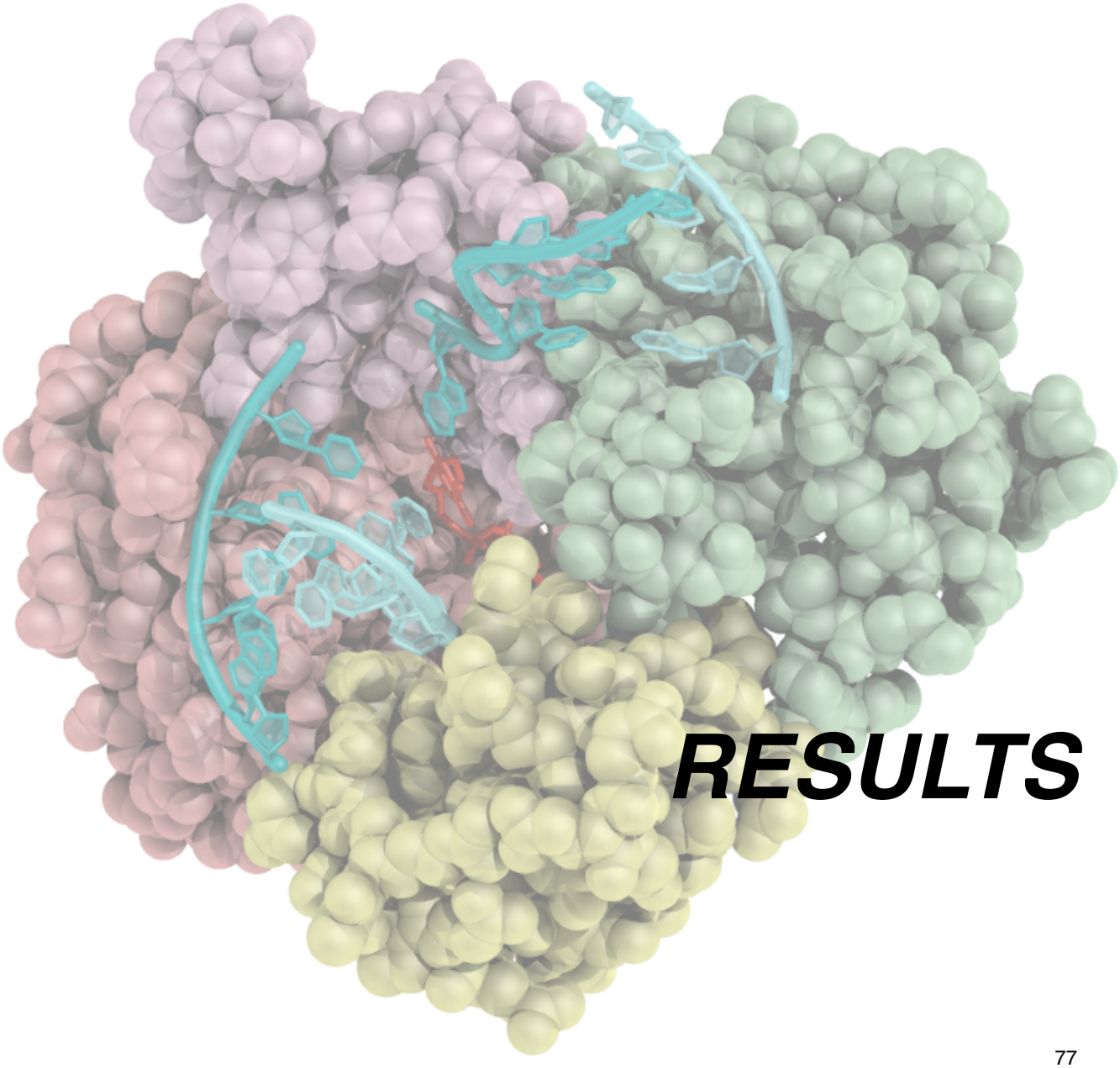
| NAME | SEQUENCE (5' -> 3') *(Table 1)* |
|---|---|
| PolydT | TTTTTTTTTTTTTTTTTTTT |
| PolydA | AAAAAAAAAAAAAAAAAAAA |
| Sp1C | GATCACAGTGAGTAC |
| T13C | AGAAGTGTATCTCGTACTCACTGTGATC |
| T13G | AGAAGTGTATCTGGTACTCACTGTGATC |
| T13T | AGAAGTGTATCTTGTACTCACTGTGATC |
| DG1 | AGATACACTTCT |
| P15 | TCTGTGCAGGTTCTT |
| T17 | TGAAGTCCCTCTCGACGAAGAACCTGCACAGA |
| DG2 | TCGAGAGGGACTTCA |
| T18 | ACTGGCCGTCGTTCTATTGTACTCACTGTGATC |
| DG5 | AACGACGGCCAGT |
| FP-P | GGCCACGCAATGTTGACGTTTTTCGACAAGACCTCAGTAT |
| FP-T | GGCAGCTTGGATCTTGTCGAAAAACGTCAACATTCGCCTAGGCTTCGGCAATACTGAGGTCT TGTCGAAAAACGTCAACATTGCGTGGCC |
| FP-D | GCCGAAGCCTAGGCGAATGTTGACGTTTTTCGACAAGATCCAAGCTGCC |
| 1D-NHEJ | GGGAGGGAGGG |
| 2D-NHEJ | GGGACGTGAGTGC |
| 1A | CCCTCCCTCCCA |
| 1C | CCCTCCCTCCCC |
| 1G | CCCTCCCTCCCG |
| 1T | CCCTCCCTCCCT |
| 2A | GCACTCACGTCCCA |
| 2C | GCACTCACGTCCCC |
| 2G | GCACTCACGTCCCG |
| 2T | GCACTCACGTCCCT |
| 1AC | CCCTCCCTCCCCA |
| 1AAC | CCCTCCCTCCCCAA |
| 1AAAC | CCCTCCCTCCCCAAA |
| 1TG | CCCTCCCTCCCGT |

| NAME | SEQUENCE (5' -> 3') *(Table 1)* |
|---|---|
| 1TTG | CCCTCCCTCCCGTT |
| 1TTTG | CCCTCCCTCCCGTTT |
| 2AC | GCACTCACGTCCCCA |
| 2TG | GCACTCACGTCCCGT |
| D1 | CGGAGGGAGGG |
| D2 | GGGACGTGAGTGCGCG |
| D3-A | CCCTCCCTCCGAGGC |
| D3-C | CCCTCCCTCCGCGGC |
| D3-G | CCCTCCCTCCGGGGC |
| D3-T | CCCTCCCTCCGTGGC |
| D4-A | CGCGCACTCACGTCCCAGCC |
| D4-C | CGCGCACTCACGTCCCCGCC |
| D4-G | CGCGCACTCACGTCCCGGCC |
| D4-T | CGCGCACTCACGTCCCTGCC |
| 1-GCC | CCCTCCCTCCCGCC |
| 2-CGG | GCACTCACGTCCCCGG |
| D-nick-C | GGGAGGGAGGGC |
| D-nick-G | GGGAGGGAGGGG |
| D4-AC | CGCGCACTCACGTCCCCAGCC |
| D4-CC | CGCGCACTCACGTCCCCCGCC |
| D4-GC | CGCGCACTCACGTCCCCGGCC |
| D4-TC | CGCGCACTCACGTCCCCTGCC |
| D4-CA | CGCGCACTCACGTCCCACGCC |
| D4-CG | CGCGCACTCACGTCCCGCGCC |
| D4-CT | CGCGCACTCACGTCCCTCGCC |
| D3-MM | CCCTCCCTCCGCGAC |
| D3-BB1 | CCCTCCCTCCGCGGAC |
| D3-BB2 | CCCTCCCTCCGCGAGC |
| D3-FLAP | CCCTCCCTCCGCGGA |
| D4-MM | CGCGCACTCACGTCCCCGAC |

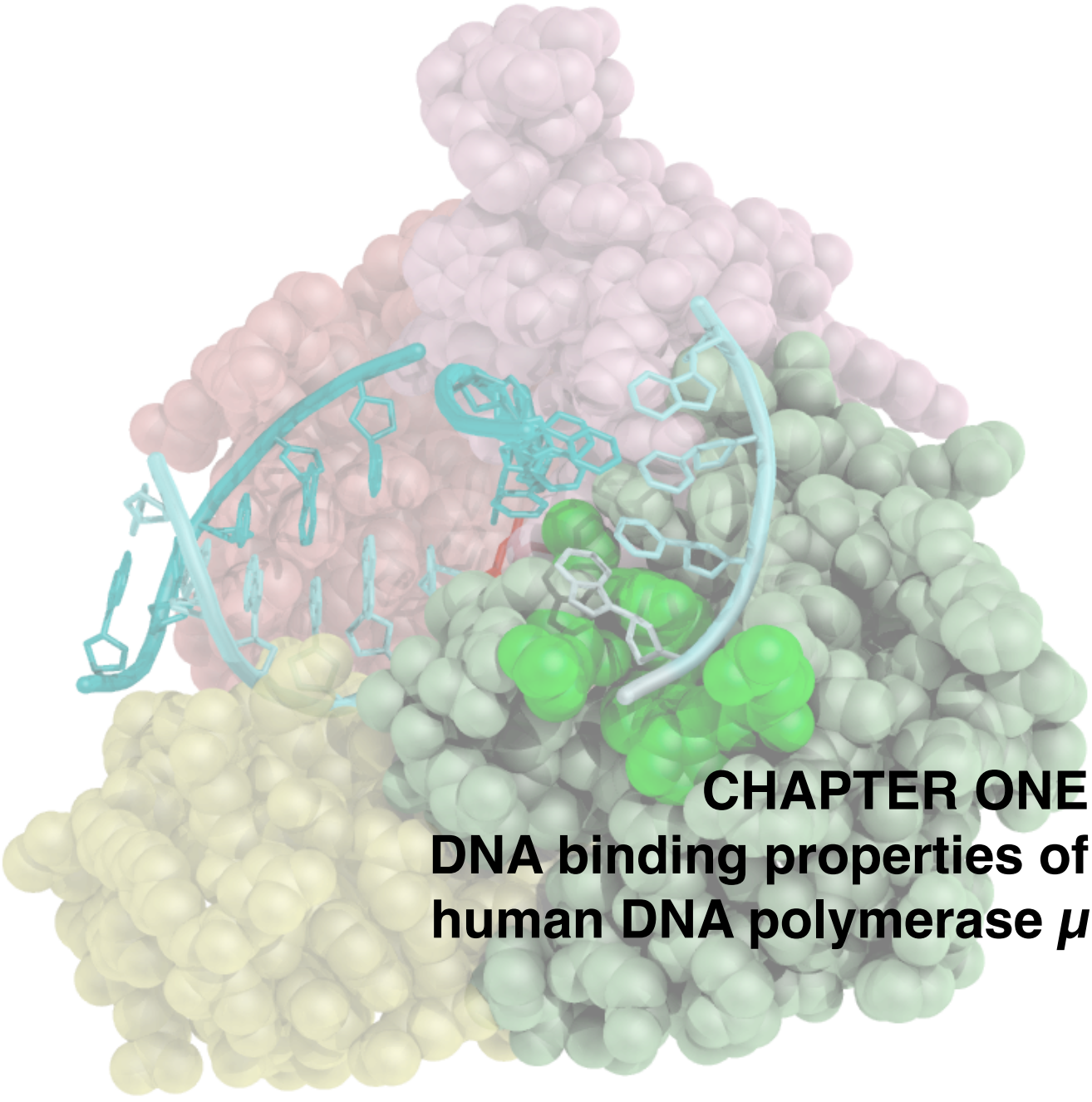
| NAME | SEQUENCE (5' -> 3') *(Table 1)* |
|---|---|
| D4-BB1 | CGCGCACTCACGTCCCCGCAC |
| D4-BB2 | CGCGCACTCACGTCCCCGACC |
| D4-FLAP | CGCGCACTCACGTCCCCGCA |
| 1GC | CCCTCCCTCCCCG |
| 2CG | GCACTCACGTCCCCG |
| 2C8 | GCACTCACGTCCCC8 |
| pBER | CTGCAGCTGATGCGCC |
| T4 | GTACCCGGGGATCCGTACGGGCGCATCAGCTGCAG |
| T8 | GTACCCGGGGATCCGTAC8GGCGCATCAGCTGCAG |
| dBER | GTACGGATCCCCGGGTAC |
| 1CG | CCCTCCCTCCGC |
| 2GG | GCACTCACGTCCCGG |
| 28G | GCACTCACGTCC8G |
| 1AG | CCCTCCCTCCCGA |
| TdT-B | CGCAAGTCAGCGCTACGGG |
| TdT-1T | CGCAAGTCAGCGCTACGGGT |
| TdT-2T | CGCAAGTCAGCGCTACGGGTT |
| TdT-3T | CGCAAGTCAGCGCTACGGGTTT |
| TdT-D | CCCGTAGCGCTGACTTGCG |
| GG | GG |
| GGG | GGG |

| NAME | SEQUENCE (5' -> 3') *(Table 2)* |
|---|---|
| S12A-F | CGAGCGCGGGTCGGGGCCCCTAGCGGCGAT |
| S12A-R | ATCGCCGCTAGGGGCCCCGACCCGCGCTCG |
| S12E-F | CGAGCGCGGGTCGGGGAGCCTAGCGGCGAT |
| S12E-R | ATCGCCGCTAGGCTCCCCGACCCGCGCTCG |
| T21A-F | GCGTGCCAGAGAAAGGCACCGCTTAAATGTG |
| T21A-R | CACATTTAAGCGGTGCCTTTCTCTGGCACGC |
| T21E-F | GCGTGCCAGAGAAAGGAACCGCTTAAATGTG |
| T21E-R | CACATTTAAGCGGTTCCTTTCTCTGGCACGC |
| Y141F-F | TGGATGCCTGCCTTTGCCTGCCAGCGC |
| Y141F-R | GCGCTGGCAGGCAAAGGCAGGCATCCA |
| Y141S-F | TGGATGCCTGCCTCTGCCTGCCAGCGC |
| Y141S-R | GCGCTGGCAGGCAGAGGCAGGCATCCA |
| C143G-F | CCTGCCTATGCCGGCCAGCGCCCTACG |
| C143G-R | CGTAGGGCGCTGGCCGGCATAGGCAGG |
| R145A-F | TATGCCTGCCAGGCCCCTACGCCCCTC |
| R145A-R | GAGGGGCGTAGGGGCCTGGCAGGCATA |
| R145K-F | TATGCCTGCCAGAAACCTACGCCCCTC |
| R145K-R | GAGGGGCGTAGGTTTCTGGCAGGCATA |
| N153G-F | CTCACACACCACGGCACTGGCCTCTCC |
| N153G-R | GGAGAGGCCAGTGCCGTGGTGTGTGAG |
| K249A-F | GGGGTCGGTGTGGCGACTGCTGACCGG |
| K249A-R | CCGGTCAGCAGTCGCCACACCGACCCC |
| R253A-F | AAGACTGCTGACGCGTGGTACCGGGAA |
| R253A-R | TTCCCGGTACCACGCGTCAGCAGTCTT |
| H329G-F | AAGTTGCAGGGCGGAGACGTGGACTTC |
| H329G-R | GAAGTCCACGTCTCCGCCCTGCAACTT |
| S372A-F | GCACAGCTGCTGTGAGGCCCCTACCCGCCTG |
| S372A-R | CAGGCGGGTAGGGGCCTCACAGCAGCTGTGC |
| S372E-F | GCACAGCTGCTGTGAGGAACCTACCCGCCTG |
| S372E-R | CAGGCGGGTAGGTTCCTCACAGCAGCTGTGC |

| NAME | SEQUENCE (5' -> 3') *(Table 2)* |
|---|---|
| F385G-F | CACATGGACGCTGGTGAGAGAAGTTTC |
| F385G-R | GAAACTTCTCTCACCAGCGTCCATGTG |
| R387A-F | GACGCTTTTGAGGCGAGTTTCTGCATT |
| R387A-R | AATGCAGAAACTCGCCTCAAAAGCGTC |
| R387K-F | GACGCTTTTGAGAAGAGTTTCTGCATT |
| R387K-R | AATGCAGAAACTCTTCTCAAAAGCGTC |
| F389L-F | TTTGAGAGAAGTTTATGCATTTTCCGC |
| F389L-R | GCGGAAAATGCATAAACTTCTCTCAAA |
| R416A-F | TGGAAGGCCGTGGCAGTGGACTTGGTA |
| R416A-R | TACCAAGTCCACTGCCACGGCCTTCCA |
| R442A-F | AAGCTTTTCCAGGCGGAGCTGCGCCGC |
| R442A-R | GCGGCGCAGCTCCGCCTGGAAAAGCTT |
| R442K-F | AAGCTTTTCCAGAAGGAGCTGCGCCGC |
| R442K-R | GCGGCGCAGCTCCTTCTGGAAAAGCTT |
| R445A-F | CAGCGGGAGCTGGCGCGCTTCAGCCGG |
| R445A-R | CCGGCTGAAGCGCGCCAGCTCCCGCTG |
| R449G-F | CGCCGCTTCAGCGGGAAGGAGAAGGGC |
| R449G-R | GCCCTTCTCCTTCCCGCTGAAGCGGCG |
| H459G-F | TGGCTGAACAGCGGTGGGCTGTTTGAC |
| H459G-R | GTCAAACAGCCCACCGCTGTTCAGCCA |



# RESULTS



# CHAPTER ONE
# DNA binding properties of human DNA polymerase *μ*

Studying the DNA binding properties of a DNA polymerase is of central importance in order to elucidate its role in different cellular processes. Both quantitative and qualitative analysis of binding to different DNA polymerization substrates can reflect intrinsic properties as processivity *versus* distributivity, and intrinsic preferences for DNA repair *versus* DNA replication substrates.

For this study we used electro-mobility shift assays (EMSA), which allow measurement of the formation of stable DNA/protein complexes either in the absence or in the presence of nucleotide and metal ions, which are necessary for catalysis. Different DNA substrates may be used in these assays, i.e. to mimick the structures generated in various DNA repair processes. In this study, a *footprinting* assay for X family polymerases was also developed; this technique allowed further study of qualitative binding differences caused by the addition of other substrates or by the lack of certain domains in the protein.

**1. DNA binding properties of human family X polymerases.**

The DNA substrates that are most commonly used for the DNAseI footprint technique are double-stranded, blunt-ended DNA fragments. However, these are not valid substrates for DNA polymerases, as they normally require ssDNA portions, recognized as the template strand. DNA polymerases from the X family are even more exquisite for binding, being DNA gaps their preferred substrate. Thus, we decided to use a 1 nucleotide-gap containing substrate, produced by hybridization of three oligonucleotides: template (FP-T, 90mer; radioactively labelled at its 5'-end), primer (FP-P, 40mer) and downstream (FP-D, 49mer), as shown in figure 18A. As a positive control of hybridization of the three oligonucleotides, the gapped substrate conforms two consensus binding boxes for Spo0A (*Bacillus subtillis* sporulation factor; Castilla-Llorente et al., 2006), located at both the primer and downstream sides of the gap. As expected, strong binding of this control protein to the DNA substrate produced a large retardation band using EMSA (Fig. 18B, left panel), and two clear footprints indicating that, indeed, Spo0A was bound to the two specific boxes (Fig. 18C, first panel). The DNA binding affinity of the four members of the human family X to this 1-nt gapped substrate was tested by EMSA. Both Polß (7,5 $\mu$M) and Pol$\mu$ (1,5 $\mu$M) formed stable DNA/protein complexes, while to obtain a similarly stable complex with Polλ the amount of polymerase needed was 5x higher (Fig. 18B, right panel). In agreement with these results, both Polß and Pol$\mu$ produced a clear footprint on this DNA substrate in the assay shown in figure 18C, second panel, lanes 2 and 4, being the protection caused by equal amounts of Polλ visibly weaker (Fig. 18C, second panel, lane 3). These protections are centered around the templating base, at position 41. As expected, TdT showed no footprint at all on this substrate (Fig. 18C, second panel, lane 5), since this polymerase does not bind template-containing substrates.

When comparing the footprint of Polß with the crystal structures available of Polß bound to a gapped DNA substrate, we found a complete correspondence between the number of nucleotides (9 nt) protected in our assays (from position 36 to 44, both included: 5 nt upstream the gap/primer side, 3 nt downstream, and the templating base) and the number of base-pairs contacted by the polymerase in the 3D structure, as shown in figure 18D. In the case of Pol$\mu$, a similar protection was observed from position 36 to 44, but the footprint was further extended by about 7 nucleotides (up to position 51) towards the downstream part of the substrate. This extra protection is interpreted to be due to the presence of a BRCT domain in Pol$\mu$ (Dominguez et al., 2000), absent in Polß, as will be further discussed in Chapter 2.

**2. Importance of the 5'-phosphate flanking a gap.**

Previous studies with Polß revealed that its ability to perform processive polymerization on a gapped DNA substrate of several nucleotides depends on the presence of a 5'-phosphate group flanking the gap (Singhal and Wilson, 1993). This 5'-P does have a positive influence on the catalytic efficiency of Polß, as well as on the level of fidelity during synthesis, since both parameters are significantly increased when the 5'-P is present (Chagovetz et al., 1997). These effects are promoted in part by an increased binding to DNA, mediated by the specific interaction of this 5'-P group with a positively charged region of the 8 kDa domain in Polß, as shown in figure 19A. Such an interaction, absent in

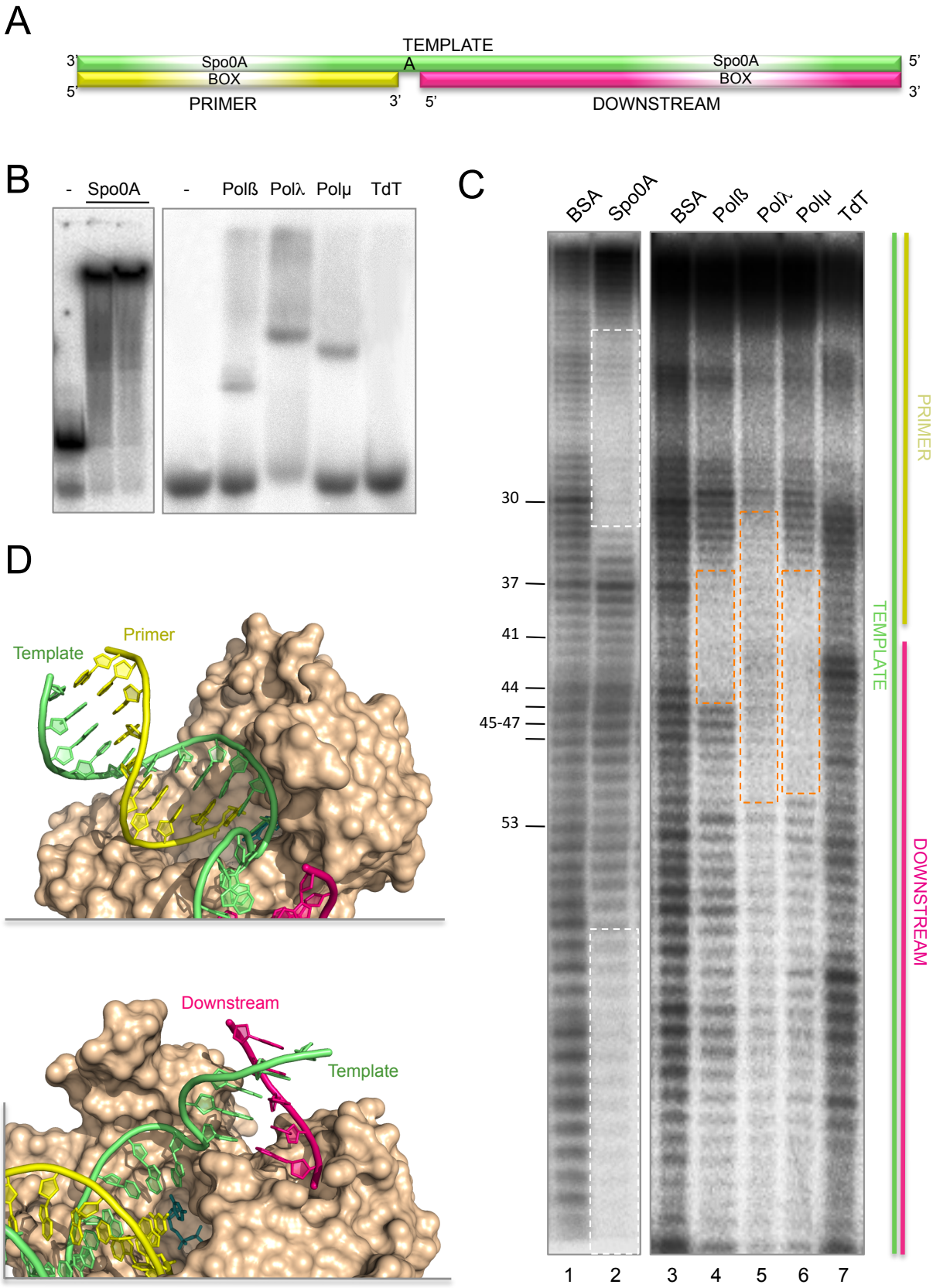



**Figure 18. DNA Footprint of X Family Polymerases.**
A) Scheme of the substrates used for the footprinting assays. To produce this substrate the oligonucleotides FP-T (template), FP-P (primer) and FP-D (downstream) were hibridized. B) DNA binding affinity of the indicated proteins (0.5 $\mu$M) was assayed as described in Materials and Methods, using the footprinting substrate radioactivelly labeled at the 5' end of the template strand. Gel was dried and the labeled fragments detected by autoradiography. C) Footprinting assay of the control protein Spo0A (1$\mu$g) or each of the members of the X family (Polß, 5 $\mu$g; Pol$\mu$, 1.5 $\mu$g; Polλ, 10 $\mu$g, TdT, 10 $\mu$g) was conducted as described in Materials and Methods. 10 $\mu$g of BSA were added to the control lane. D) Views of the crystal structure of Polß ternary complex (1BPZ) in which the protein is shown in surface, the DNA substrate is shown in sticks, following the same coloring as in A),and the incoming dNTP is shown in blue.

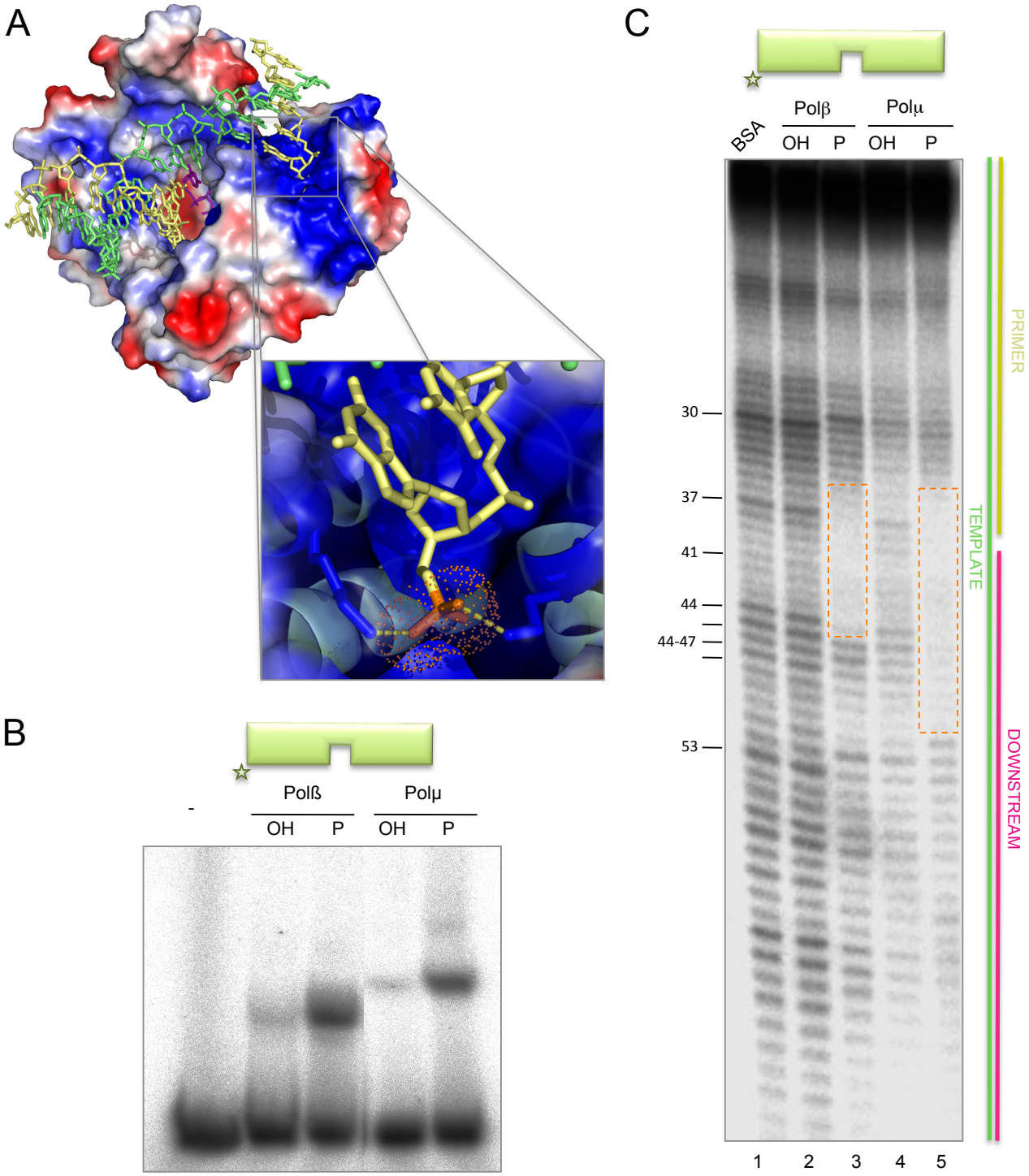



**Figure 19. Importance of the 5'-P group for Polß and Pol*μ* binding to the DNA substrate.**
A) Crystl structure of Polß showing the electrostatic surface of the polymerase and the DNA substrate in sticks (template strand, green; primer and downstream strands, yellow). The inset shows a zoomed-in view of the downstream 5'-P group binding pocket and its contacts with the positively charged residues that form the pocket. B) EMSA of Polß and Pol*μ* (both at 100 nM) using a gapped DNA substrate formed by the oligonucleotides Sp1C (labeled at the 5' end), T13C and Dg1, tha latter either having (P) of lacking (OH) a 5'-P group. Gel was dried and the labeled fragments detected by autoradiography. C) Footprinting assay of Polß (5 *μ*g) or Pol*μ* (1.5 *μ*g) with a gapped substrate that either contains (P) or lacks (OH) a 5'-P group in the downstream strand. 10 *μ*g of BSA were added to the control lane. Gel was dried and the labeled fragments detected by autoradiography.

replicative polymerases, is very specific as the 5'-P hallmarks the end point of polymerization, i.e. the limit of the DNA gap which Polß is filling in. The amino acid sequence and structural similarities between Polß and Pol*μ*, both having a conserved 8 kDa domain, prompted us to characterize if 5'-P binding also contributes to Pol*μ* DNA binding

properties in order to further discuss its hypothetical role in different cellular processes of DNA synthesis.
We evaluated the capacity of Pol*μ* to bind gapped substrates which, when indicated, may contain a 5'-P group in the downstream strand, thus flanking the gap. EMSA showed a strong increase in the binding affinity of both Polß and Pol*μ* to the 5'-P containing substrate (**Fig**. 19B). Footprinting assays corroborated this result (Fig. 19C; compare lanes 2&3 for Polß and lanes 4&5 for Pol*μ*). In fact, the footprint by Polß and Pol*μ* is hardly visible on the gapped substrate lacking the 5'-P. Thus, both polymerases display a higher affinity for gap-containing DNA substrates bearing a 5'-P, similar to those generated during several DNA repair pathways.

**3. Ternary complex formation**

Analysis of the available 3D information shows structural differences in Polß when forming binary (E:DNA) *versus* ternary (E:DNA:NTP) complexes. Moreover, previous data from our laboratory had shown that both Polλ binary and ternary complexes can be detected and distinguished by EMSA (Picher, 2007). To compare Polß and Pol*μ* binary and ternary complexes, we decided to use our footprinting assay on a 1nt-gapped DNA either in the absence (binary) or in the presence (ternary) of a deoxynucleotide (complementary or non-complementary to the templating base).
In the case of Polß we can see that upon addition of the complementary deoxynucleotide (dT), but not the non-complementary (dA), the footprint of the polymerase was more obvious (see band at position 37) and expanded further into the primer zone (Fig. 20B, compare lanes 2-4), indicative of an adjustment of the polymerase conformation, allowing a stronger binding to the primer region. Upon addition of either deoxynucleotide a new hypersensitivity appears at position 29. It should be noted that there was an initial hypersensitivity at position 30, i.e., frequent cuts or *nicks* are produced by DNAseI at this position in the DNA substrate. These nicks are an ideal substrate for Polß, since they mimick the intermediate substrates of the BER pathway in which this polymerase is implicated. Thus, Polß would insert a deoxynucleotide on this nicked substrate, leading to the formation of a +1 product that appears at position 29. Changes at other bands can be also explained as polymerization events, not related to the specific footprint centered at the gap (position 41).
Also in the case of Pol*μ*, the area protected by the polymerase was slightly extended only upon addition of the correct deoxynucleotide (Fig. 20B, lane 7), indicating that the complementary deoxynucleotide promotes an adjustment that allows a better binding to the primer area. On the other hand, we can see the appearance of the hypersensitivity at position 29, but unlike what happened with Polß, the hypersensitivity only changed from position 30 to 29 when adding the complementary nucleotide (Fig. 20B, lane 7), and not with the non-complementary one (lane 6).
To confirm that the extension of the footprint was in fact due to the interaction with the nucleotide, we used two mutants of Pol*μ*: H329G (eliminating the function of a residue involved in nucleotide binding; Fig 3A), and R387A (mutating not a nucleotide ligand, but a dual [template/primer] DNA ligand, thus allowing ternary complex formation; Fig. 20A). As expected, the footprint obtained with mutant H329G does not expand with the addition of nucleotide, either correct or incorrect (Fig. 20C, lanes 3 and 4), while the footprint produced by mutant R387A, which itself is able to bind the nucleotide, is clearly elongated upon addition of the correct dTTP (Fig. 20C, lane 8). These mutants do not display the downstream expansion of the footprint probably corresponding to the BRCT domain, and this could be due to their inability to form a correct binary complex with the DNA substrate. That defect was suppressed only in mutant R387A upon addition of the correct (dT) nucleotide. The different phenotypes displayed by these mutants will be discussed in further detail in Chapters 4 and 5.

**4. Intrinsic Pol*μ* binding to NHEJ substrates.**

Pol*μ* has been implicated in the NHEJ pathway of DSB repair (Ramsden, 2011). This DNA repair mechanism implies bridging of two DNA termini, that can be either blunt or 5'-/3'-protruding. In the case of 5'-protruding ends, DNA synthesis can occur in a step prior to end-bridging due to the simultaneous presence of a recessive 3'-hydroxyl group and a template strand, both being requisites for most DNA polymerases, and the resulting blunt ends could be directly ligated. Conversely, polymerization on a 3'-

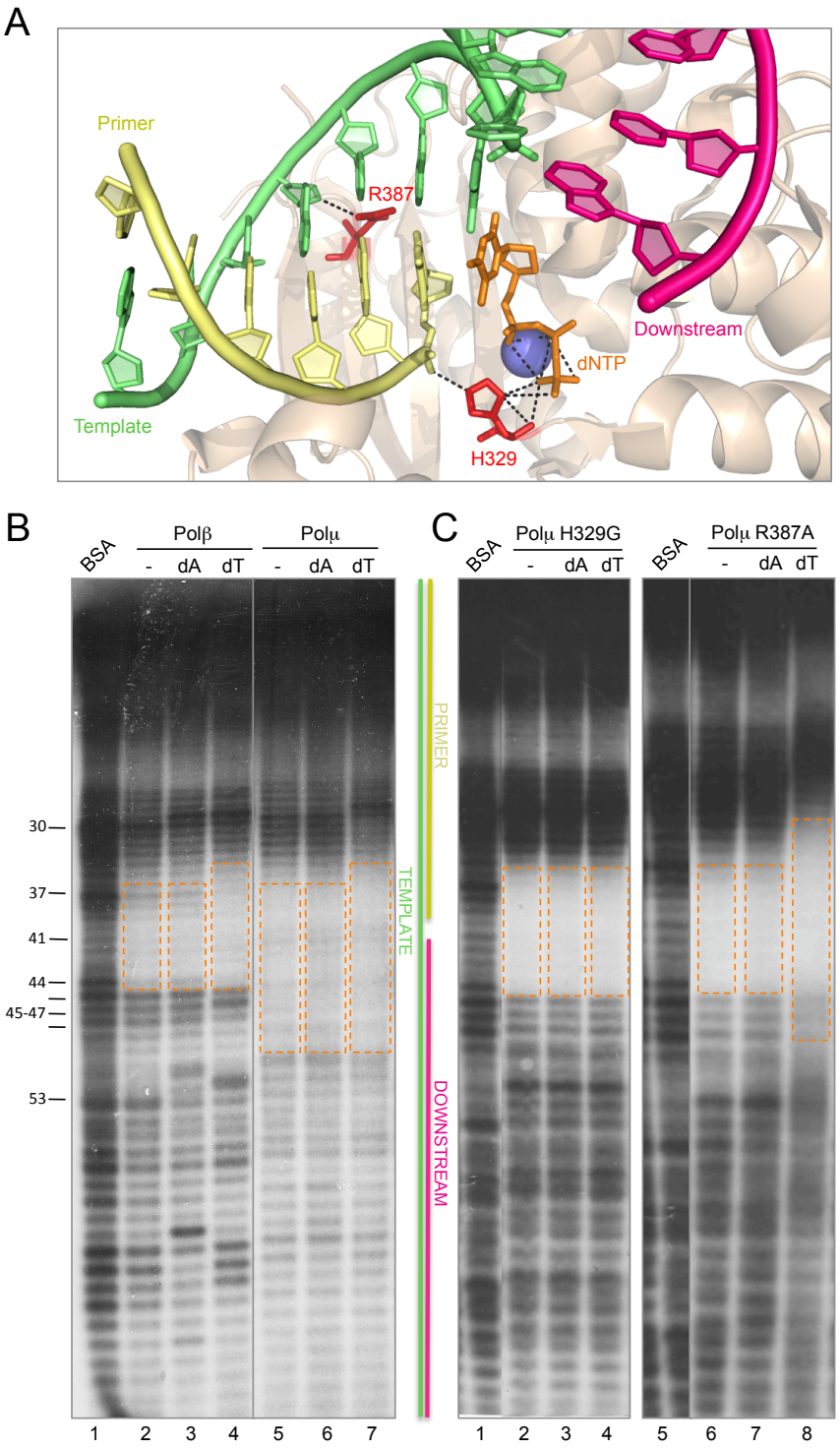



**Figure 20. Formation of the ternary complex.**
A) Structure of Pol*μ* (shown in wheat-colored ribbons) ternary complex (2IHM), in which the DNA substrate and incoming nucleotide are shown in sticks with the following colors: dNTP, orange; template strad, green; primer strand, yellow; downstream strand, dark pink. Selected residues are shown in red sticks. B) Footprinting assay of the wild-type Polß (5 *μ*g) and Pol*μ* (1.5 *μ*g). When indicated, 100 *μ*M dATP or dTTP were added, together with 2.5 mM $MgCl_2$. The DNA substrate used, formed by hybridizing the olignucleotides FP-T (template, labeled at its 5' end), FP-P (primer) and FP-D (downstream) always contains a phosphate group at the 5'end of the downstream strand. Gel was dried and the labeled fragments detected by autoradiography. C) Footprintg assay of Pol*μ* mutants H329G and R387K (1.5 *μ*g) were carried aout as described in B).

protruding end requires either a terminal transferase activity or the previous connection of the two ends to provide a template strand. The ability to bind this kind of 3'-protruding DNA substrates by a DNA polymerase would imply an advantage for filling in the gaps formed during the end-joining process. Binding to these molecules can occur in two ways, since the DNA polymerase could recognize the 3'-protrusion either as the 3´-primer *terminus* or as a template strand. To elucidate the binding mode of Pol*μ* to these substrates, we tested 3'-protruding molecules either bearing or lacking a 5'-P group in the recessive strand. In order to minimize any possible self-pairing, allowed by molecules containing very long single stranded portions, we decided to use a new set of molecules with very

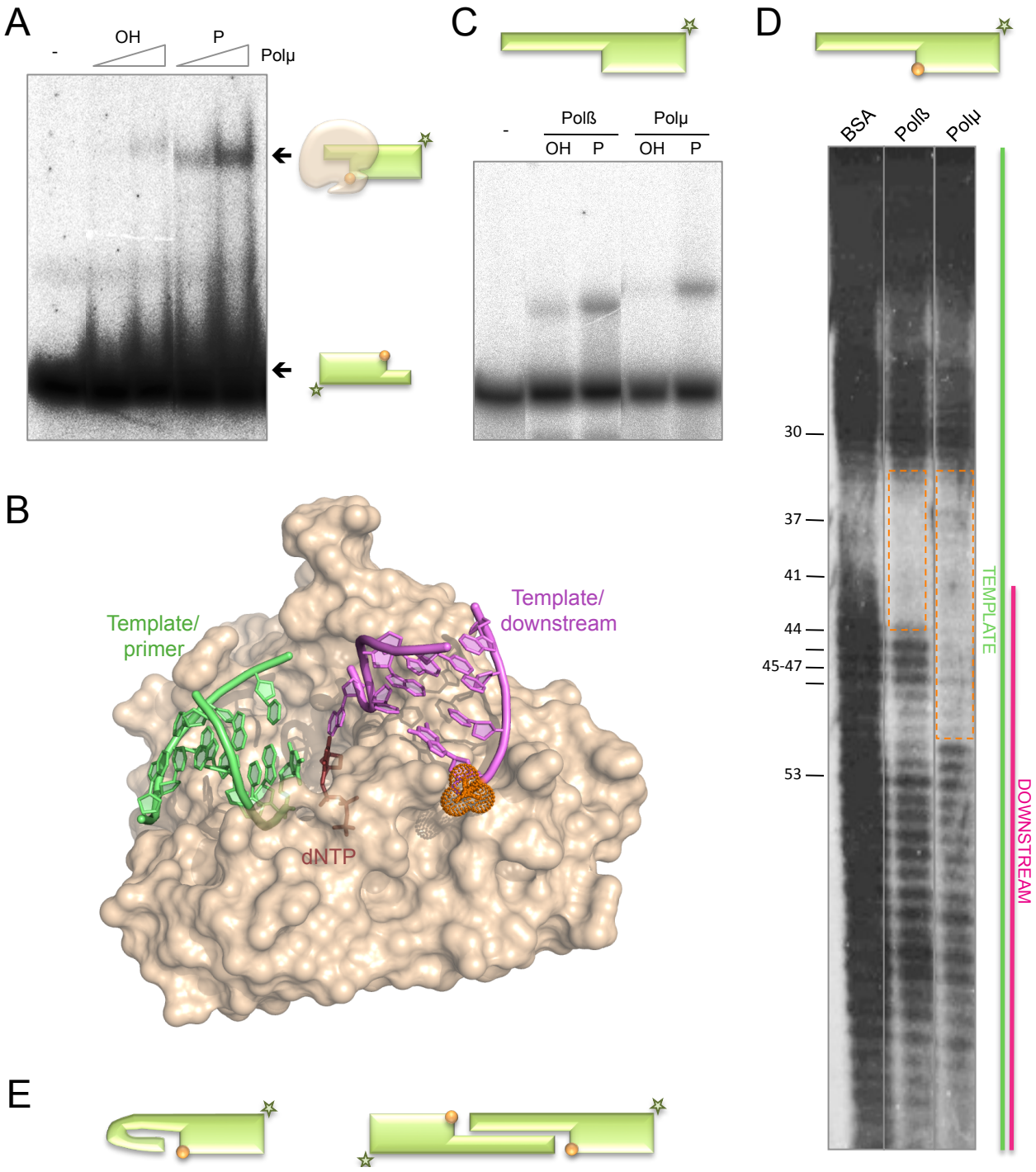
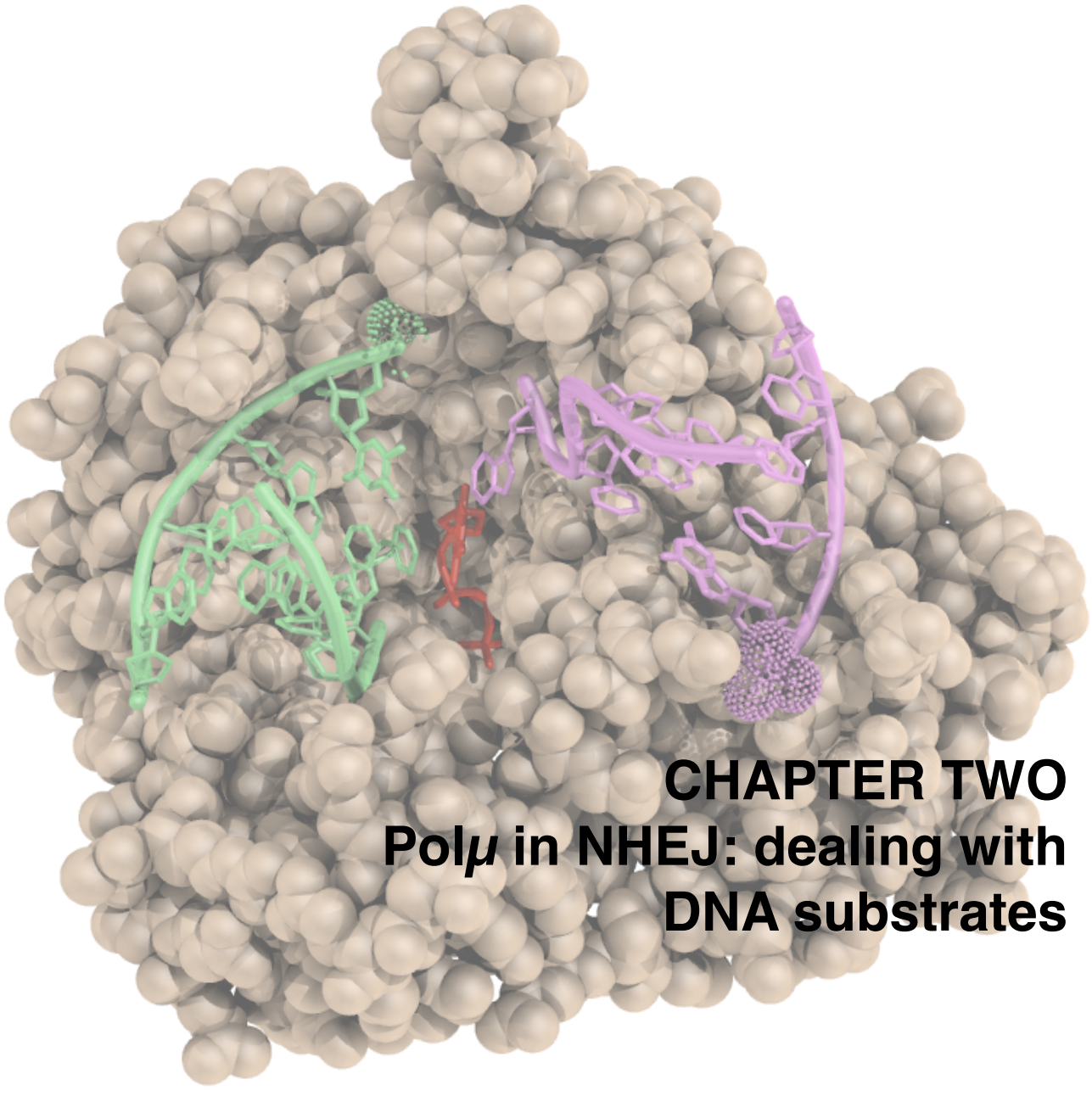



**Figure 21. Intrinsic Pol*μ* binding to template/downstream substrates.**
A) EMSA of Polß and Pol*μ* (600 nM each) using T/D substrates (formed by hybridization of T13C and DG1 oligonucleotides) either lacking (OH) or having (P) a phosphate group at the 5' end of the downstream strand. The schemes show the free substrate (green with a yellow ball depicting the 5'-P group) or the polymerase bound to the substrate. B) Surface representation of Pol*μ* bound to two DNA ends (shown in mauve -T/D- and green -T/P) and incoming nucleotide (maroon). The 5'-P group of the downstream strand is shown as orange dots. C) EMSA of Polß or Pol*μ* (200 nM each) were carried out using a T/D substrate formed by the hybridization of FP-T and FP-D oligonucleotides, either having (P) or lacking (OH) a 5'-P group. After electrophoresis, the gel was dried and the labeled fragments were detected by autoradiography. D) Footprinting assaya of Polß (5 *μ*g) and Pol*μ* (1.5 *μ*g) using the same substrate as in C), always containing the 5'-P group. After electrophoresis, the gel was dried and the labeled fragments were detected by autoradiography. E) Schematic representation of the two possible arrangements of the substrates mediated by the polymerase when contronted with a T/D with a long 3'-protrusion.

short (3 nt) 3'-protruding ends. By using EMSA with these substrates we concluded that Pol*μ* is able to form a stable complex when binding 5'-P-containing 3'-protruding substrates (Fig. 21A). Considering the previous results using gapped substrates (Fig. 19), and the recognition of the 5´ phosphate by the 8 kDa domain, it is very likely that the presence of the 5´-P orients binding, thus selecting the 3'-protrusion as a template strand (mauve colored DNA substrate in figure 21B). When the phosphate is not present, a much weaker binding to the 3'-protrusion is occurring, likely oriented as a primer strand (green colored DNA substrate in figure 21B). This specific binding, reinforced by the 5'-P, would help Pol*μ* to initiate bridging of two DNA ends (a 3'-protrusion is firstly recognized as the template-providing end, and the 3'-protrusion of a second end must be recognized as primer) during the NHEJ reaction (Fig. 21B).

Given the possible implication of Pol*μ* in the NHEJ repair pathway, it was necessary to obtain further qualitative insights explaining how this polymerase binds 3'-protruding DNA ends as NHEJ intermediates, and for this we decided to apply our footprinting assay on a hybrid formed by only the template and 5'-P containing downstream oligonucleotides. As a control, we used this footprinting substrate in EMSA (Fig. 21C), and as expected, both Polß and Pol*μ* showed greater affinity for a substrate with a 5'-phosphate group in the recessive strand. Footprinting assays showed that both polymerases strongly protected the DNA area around the 5'-recessive end (Fig. 21D). Moreover, it should be noted that the length of the protections caused on this substrate by the polymerases fits exactly with the number of base pairs covered on gap-containing substrates. This observation, together with the appearance of DNAseI cuts along the single stranded template (DNAseI preferentially cuts double stranded DNA), led us to think that the upstream area is no longer single-, but double-stranded. As a result of this, in addition to bind the double-stranded downstream region, both DNA polymerases would bind/stabilize the pseudo-primer region which is double-stranded, formed either by promoting the search for microhomology *in cis* (snapping-back the 3'-terminus of the same DNA molecule; see figure 21E, left scheme) or *in trans* (with the 3'-protruding end belonging to another molecule; figure 21E, right scheme).

The analysis of Pol*μ* DNA binding properties described in this section emphasizes the importance of the presence of a 5'-phosphate group, specially at 3'-protruding NHEJ substrates. A necessary further step was to evaluate the impact of such binding properties in the enzymatic activity of Pol*μ*.



# CHAPTER TWO
# Pol*μ* in NHEJ: dealing with DNA substrates

The non-homologous end-joining (NHEJ) pathway is the principal mechanism used to repair double strand breaks (DSBs) in higher eukaryotes. The two DNA ends may contain regions with a certain level of microhomology, at different distances from the 3'-terminus. A 5'➔3' resection would then be able to generate 3'-protruding ends susceptible of base-pairing. The different location of these microhomologies would require distortions in the DNA to optimize base-pairing. In the most extreme case, the two ends would not display any microhomology and would be non-complementary. The template-directed polymerases from the human X family, i.e. Polß, Polλ, and Pol*μ*, select their optimal DNA substrates on the basis of a gradient of template-dependence (maximal in Polß, moderate in Polλ, and much lower in the case of Pol*μ*; {Nick McElhinny et al., 2005). In agreement with these differences, it has been shown that Polß is not a player of the NHEJ pathway, while Polλ and Pol*μ* are the main polymerases implicated; however, they are not redundant since in the presence of NHEJ accessory factors such as Ku70/80 and XRCC4/LigaseIV, both of them can use complementary ends but only Pol*μ* is able to join non-complementary ends (Nick McElhinny et al., 2005).

**1. Family X in NHEJ: Polß, Polλ, and Pol*μ***

The three template-directed members of the family X are similarly competent when polymerizing on a small gap, as shown in figure 22A. We then tested the three polymerases Polß, Polλ, and Pol*μ*, in a NHEJ *in vitro* reaction in which the polymerase alone, in the absence of NHEJ accessory factors, is challenged to bridge two DNA ends, extending one 3'-end by copying the templating base provided by a second 3'-end. For that we used two pairs of short double-stranded DNA molecules: one pair containing a 3'-overhang of 2 nucleotides which can form one base pair (microhomology; Fig. 22B, scheme), and the other pair of DNA ends, whose 1 nucleotide 3'-protrusion (dC) provides null complementarity when confronted to each other (non-complementary; Fig. 22C, scheme). In the two cases, one DNA end (depicted in green) was labeled at the 5' terminus of the 3'-protruding strand, and therefore can be assayed for primer extension, whereas the second molecule (depicted in mauve), unlabeled, provides the templating base (dC) *in trans.* Thus, preferential extension of the primer with ddG would be indicative of an accurate first stage of NHEJ. When working with DNA substrates bearing complementary 3'-protruding ends, we observed that, as expected, Polß had no activity, while Polλ and Pol*μ* can efficiently and accurately insert the correct nucleotide dictated by the templating base provided *in trans* by the second end (Fig. 22B). On the contrary, when using DNA substrates that did not contain any microhomology, both Polß and Polλ were inactive and only Pol*μ* was able to perform an efficient and accurate NHEJ reaction (Fig. 22C).

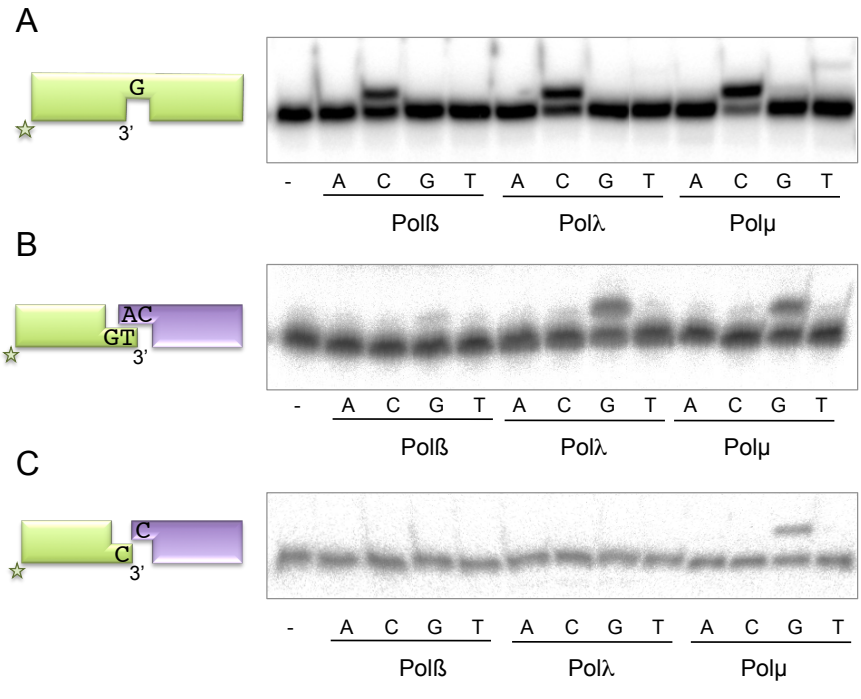



**Figure 22. End-joining activity of the template-directed X family polymerases.** A) Gap-filling activity of Polß, Polλ and Pol*μ* (25 nM each) was assayed using a substrate formed by the hybridization of the oligonucleotides SP1C, T13G and DG1-P. When indicated, 10 nM of each dNTP was added, in the presence of 2.5 mM $MgCl_2$. B) NHEJ assay of Polß (600 nM), Polλ (600 nM) and Pol*μ* (200 nM) was performed as described in Materials and Methods, using a set of compatible substrates: the labeled substrate was formed by hybridization of 1TG and 1D-NHEJ (shown in green) and the cold substrate by hybridization of 1AC and 1D-NHEJ (shown in mauve). When indicated, dNTPs were added separately at 100 *μ*M in the presence of 1 mM $MnCl_2$ for Polß and Polλ, and 2.5 mM $MgCl_2$ for Pol*μ*. After electrophoresis, the labeled fragments were detected by autoradiography. C) NHEJ reaction performed as in B), with a set of incompatible substrates in which the labeled one was formed by hybridization of 1C and 1D-NHEJ (green) and the cold one, of 2C and 2D-NHEJ (mauve). In all the cases, the substrates contain a 5'-P group at the downstream strand.

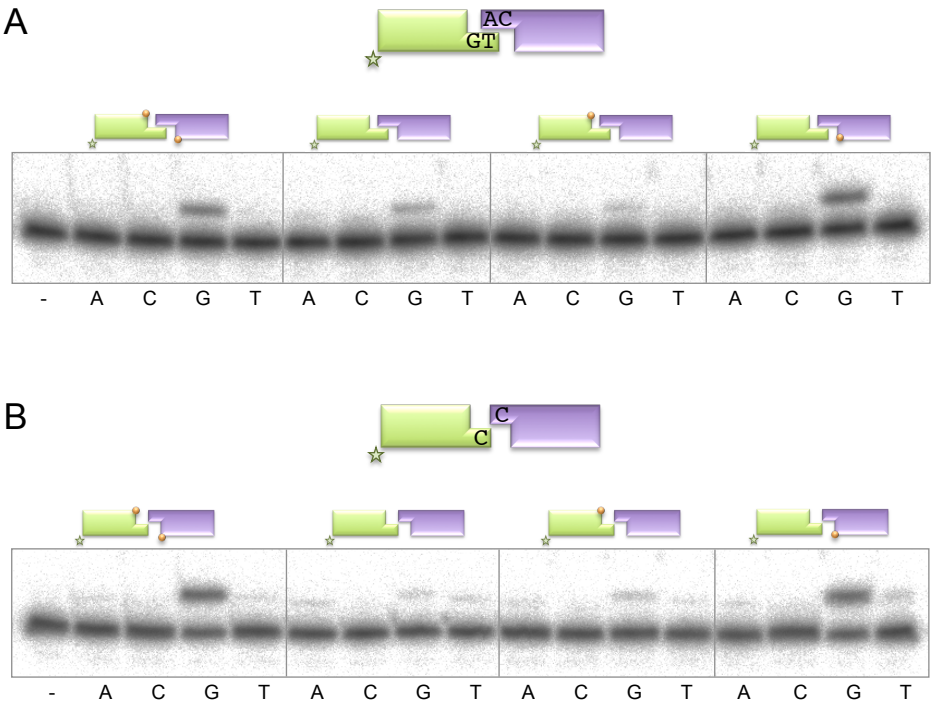



**Figure 23. Impact of the presence or absence of phosphate groups in the downstream strands of the DNA ends in the NHEJ reaction.** A) NHEJ assay of Pol*μ* (200 nM) performed as described in Materials and Methods, using a set of compatible substrates: the labeled substrate was formed by hybridization of 1TG and 1D-NHEJ (shown in green) and the cold substrate by hybridization of 1AC and 1D-NHEJ (shown in mauve). The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, dNTPs were added separately at 1 *μ*M in the presence of 2.5 mM $MgCl_2$. After electrophoresis, the labeled fragments were detected by autoradiography. B) NHEJ reaction performed as in C), with a set of incompatible substrates in which the labeled one was formed by hybridization of 1C and 1D-NHEJ (green) and the cold one, of 2C and 2D-NHEJ (mauve).

As we have shown, Pol*μ* and Polλ are intrinsically able to bridge compatible DNA ends, but only Pol*μ* can join two incompatible DNA ends, extending one 3'-end by copying the templating base provided by a second end, thus achieving a precise NHEJ reaction. A similar behaviour was shown to occur at some incompatible ends when Pol*μ* was assisted by NHEJ core factors (Davis et al., 2008). Therefore, we wanted to address a deeper analysis of the intrinsic bridging capacities of Pol*μ*, assessing both the efficiency and fidelity of the reaction on a range of DNA substrates that would cover some of the situations typically occurring during NHEJ.

## 2. Importance of 5'-P recognition for Pol*μ*-mediated NHEJ.

A general feature of DNA-dependent DNA polymerases from the X family, Polß, Polλ and Pol*μ*, is that their DNA-binding and gap-filling activity is greatly enhanced when the substrate has a 5'-P in the downstream strand, i.e. flanking the gap (Prasad et al., 1994; Garcia-Diaz et al., 2002, Chapter 1 of this PhD thesis). However, and in spite of being very early anticipated (Ruiz et al., 2001), the importance of the 5´-P for the activity of family X polymerases during NHEJ has not been directly evaluated. So, we firstly tested if the presence of this 5'-P in both or either of the two DNA ends was required for Pol*μ* bridging activity during NHEJ, by using the same kind of complementary and non-complementary pairs of molecules as in figure 22. The results showed that the 5'-P of the template-providing end (mauve) is essential for an efficient and accurate nucleotide insertion, on both compatible (Fig. 23A, first and fourth panels) and incompatible (Fig. 23B, first and fourth panels) substrates; conversely, the presence of the 5'-P in the primer-providing end (green), flanking a second gap, is unnecessary and in some cases even detrimental for this activity (Fig 23A&B, third panels). A plausible explanation, in agreement with the EMSA experiments described in the previous section, is that a recessive 5'-P only in the labeled DNA end would orient enzyme recognition in such a way that the labeled molecule would be preferentially selected to act as template (strongly competing with the cold one) instead of acting as primer. To support this idea, we tested 3 ´extension of the labelled molecules only (not including the cold ones) to evaluate the background level of terminal transferase, and also the possibility that some template-directed NHEJ reaction could be occurring among two labeled (identical) DNA ends. In the absence of a recessed 5'-P, human Pol*μ* catalyzed insertion of the four ddNTPs in a terminal transferase fashion (Fig. 24A and B, left panels),

meaning that the labelled molecule was preferentially bound as a primer (Fig. 24C, top panel). These results emphasize the change in pattern observed in figure 23 (A&B, fourth panels), where preferential extension with ddG had been dictated “in trans” by the other end, but mainly when the latter has a recessive 5’-P. If the labelled DNA ends have a recessive 5’-P, we observed different outcomes depending on the length of the overhang. In the case of the longer overhang (2 nucleotides; Fig. 24A, right panel) the observed lack of insertion of any nucleotide (i.e. terminal transferase inhibition) is indicative that Pol*μ* is now bridging two identical (non-complementary) ends, but it is not able to polymerize on them likely due to an inappropriate distance (a gap of 2 nt) between the 3´-terminus and the nucleotide biding site (Davis et al., 2008). Accordingly, the reduction shown in figure 23A (third panel) could be explained by the expected competition of the labeled DNA end as an unproductive template. In the case of the short (1 nucleotide) protrusion, a slight increase in the insertion of ddG, concomitant with a decrease in the insertion of the other three nucleotides, again supports that the 5´-P allows Pol*μ*-mediated bridging of two identical (non-complementary) molecules, one productively bound as template and other as primer (Fig. 24B, right panel; Fig. 7C, bottom panel). All these evidences indicate that when Pol*μ* encounters a DNA end bearing a recessive 5’-P, the enzyme recognizes this DNA end as a template/ downstream structure, forming a NHEJ intermediate that awaits for the arrival of the incoming primer (a second 3’-protruding DNA end) and nucleotide.

All this evidence still opens a couple of questions regarding the decisions made by Pol*μ* when both ends contain a 5’-P group: what is the schedule of events in this case? Which DNA end is bound first and in what position? To address this matter, we used a variation of the NHEJ substrates used above: in this case, the length of the two DNA end molecules was sufficiently different so that extension of both 3’-protruding strands (labeled) can be simultaneously tested; D3 is the “short” molecule

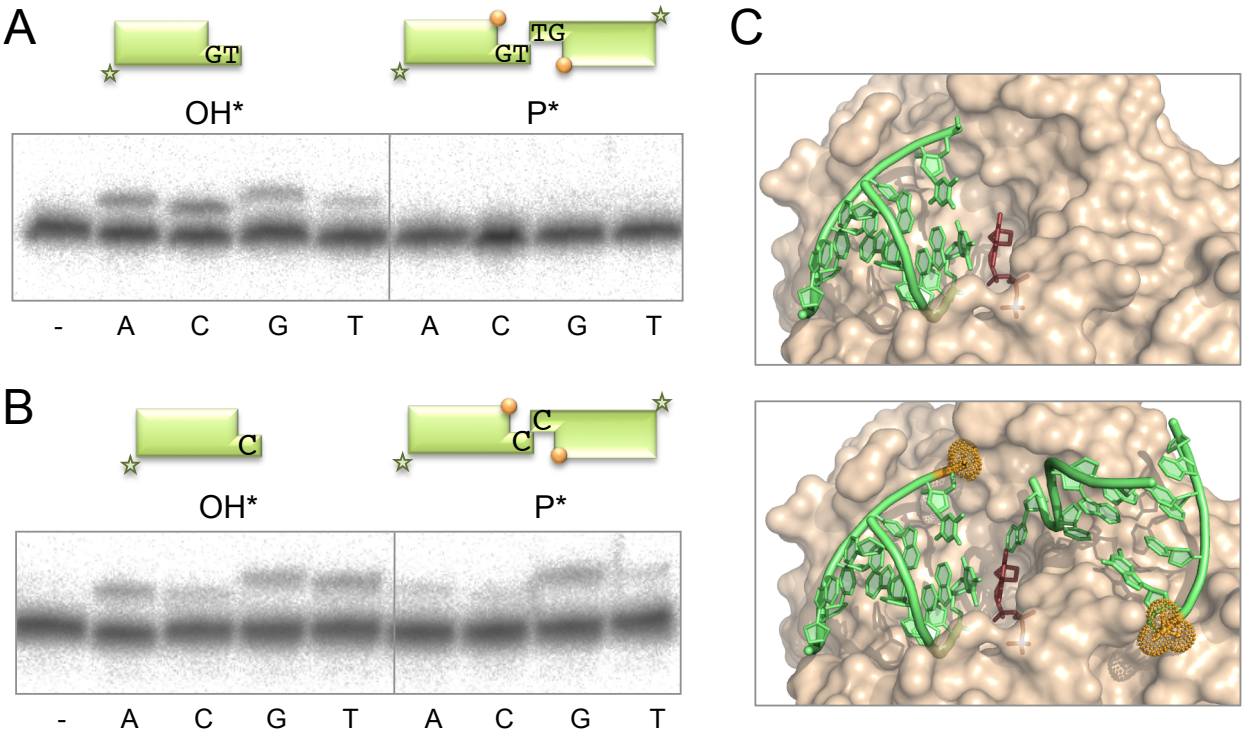



**Figure 24. The presence of a 5’-P group in the downstream strand of one DNA end promotes synapsis of two ends.**
A) NHEJ assay of Pol*μ* (200 nM) performed as described in Materials and Methods, using only one (labeled) substrate, formed by hybridization of 1TG and 1D-NHEJ (shown in green). The orange balls indicate the presence of a 5’-P group in the downstream strand of the substrate. When indicated, dNTPs were added separately at 10 *μ*M in the presence of 2.5 mM $MgCl_2$. After electrophoresis, the labeled fragments were detected by autoradiography. B) NHEJ reaction performed as in A), with one substrate formed by hybridization of 1C and 1D-NHEJ (green). C) Surface representations of Pol*μ* bound to a 1 nt 3’-protruding substrate (green) as a template primer in the absence of a 5’-P group (top panel), or as a synapsis of two DNA ends when these contain 5’-P groups (orange dots). The incoming dNTP is colored maroon.

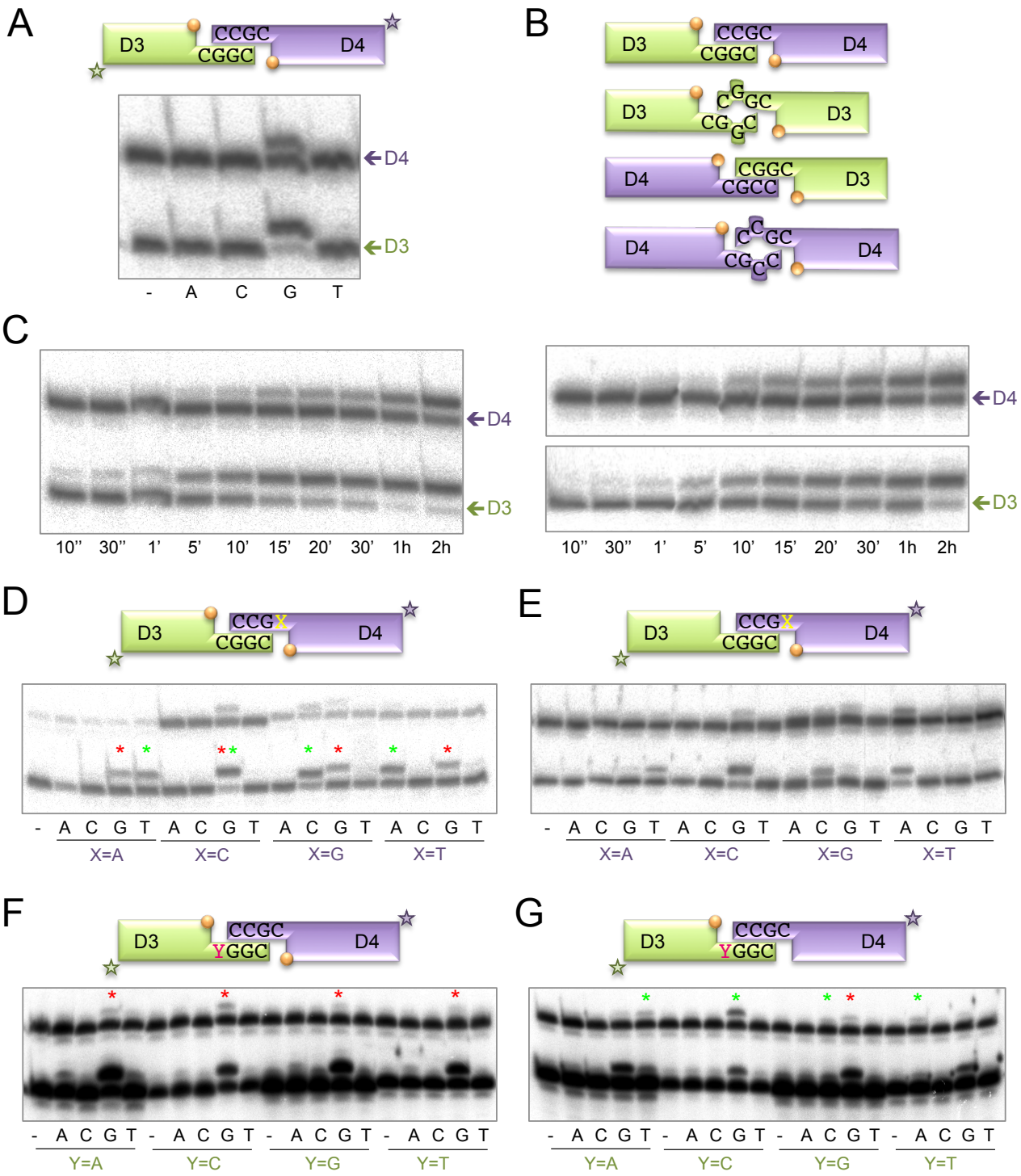



**Figure 25. The presence of a 5'-P group in a 3'-protruding DNA substrate promotes use of such as the template-providing end.** A) NHEJ reaction performed as described in Materials and Methods, with 200 nM Pol$\mu$ and a set of two labeled compatible substrates: the short substrate was formed by hybridization of D3-C and D1 (shown in green) and the long substrate by hybridization of D4-C and D2 (shown in mauve). The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, dNTPs were added separately at 1 $\mu$M in the presence of 2.5 mM $MgCl_2$. After electrophoresis, the labeled fragments were detected by autoradiography. B) Schematic representation of the possible outcomes of a reaction like the one described in A). C) NHEJ reactions were performed essentially as in A), but the period of the incubation was set to the indicated times. In the left panel, bot substrates were added. In the top right panel, only the long substrate was used, while in the bottom right anel only the short substrate was added to the reaction. D) NHEJ reactions were performed essentially as in A), but using 4 different long substrates (mauve), in which the templating base changes to any of the four possibilities. These substrates were formed by hibridization of the D2 oligonucleotide with either D4-A, D4-C, D4-G or D4-T. E) NHEJ reactions were performed essentially as in D), with the four long substrates, but this time the presence of the 5'-P group is restricted to the long substrates (mauve), while the short one (green) lacks this phosphate. F) NHEJ reactions were performed essentially as in A), but using 4 different short substrates (green), in which the templating base changes to any of the four possibilities. These substrates were formed by hibridization of the D1 oligonucleotide with either D3-A, D3-C, D3-G or D3-T. E) NHEJ reactions were performed essentially as in G), with the four short substrates, but this time the presence of the 5'-P group is restricted to the short substrates (green).

shown in green and D4 is the "long" molecule shown in mauve (Fig. 25A). These substrates display a complementarity of 3 bp (dG:dC) once joined, and thus this is also the distance between the two gaps formed. The templating base in both gaps is cytosine, and both downstream strands contain a 5'-P group. On these substrates we corroborated the NHEJ activity of Pol*μ*, this time at both sides of the break: figure 25A shows an efficient and accurate incorporation of the correct nucleotide (ddG) both in the short D3 molecule and in the long D4 substrate. However, the higher efficiency of extension of the short D3 molecule suggests that extension of the D4 end occurs as a second, delayed event. So, we decided to study the timing of the reaction in more detail, measuring the level of incorporation in a range of time points. As shown in figure 25C (left panel), the incorporation on the short primer (D3 end) occurred earlier than incorporation on the long primer (D4 end). In order to asseverate that this incorporation is occurring opposite to a templating base dictated *in trans* by the other end in each case, we repeated the same experiment but measuring Pol*μ* activity independently on each end, i.e. in the absence of the other. The results obtained indicate that Pol*μ* is able to polymerize on each end separately (Fig. 25C, right panels), either by terminal transferase or by bridging two identical molecules in a synapsis that would include distortions of the DNA strands (mismatches, etc; Fig. 25B). In the case of D3 alone, the incorporation achieved is slower than the one achieved in the presence of D4. So, when we tested the D3-D4 perfect synapsis of two different and complementary molecules (Fig. 25B, top scheme), some D3-D3 imperfect joining (Fig. 25B, second scheme) can occur. On the other hand, in the case of D4 there is no difference in the activity measured in the presence or absence of D3, suggesting that the D4-D4 imperfect synapsis would be even favored over D4-D3 (Fig. 25B, bottom).

To test the level of insertion on the D3 substrate that is template-directed *in trans* by a D4 molecule, in comparison to that arising from the D3:D3 synapsis, we used a set of D4 molecules in which the templating base (position X in figure 25D&E) is any of the four possible options. The results showed that

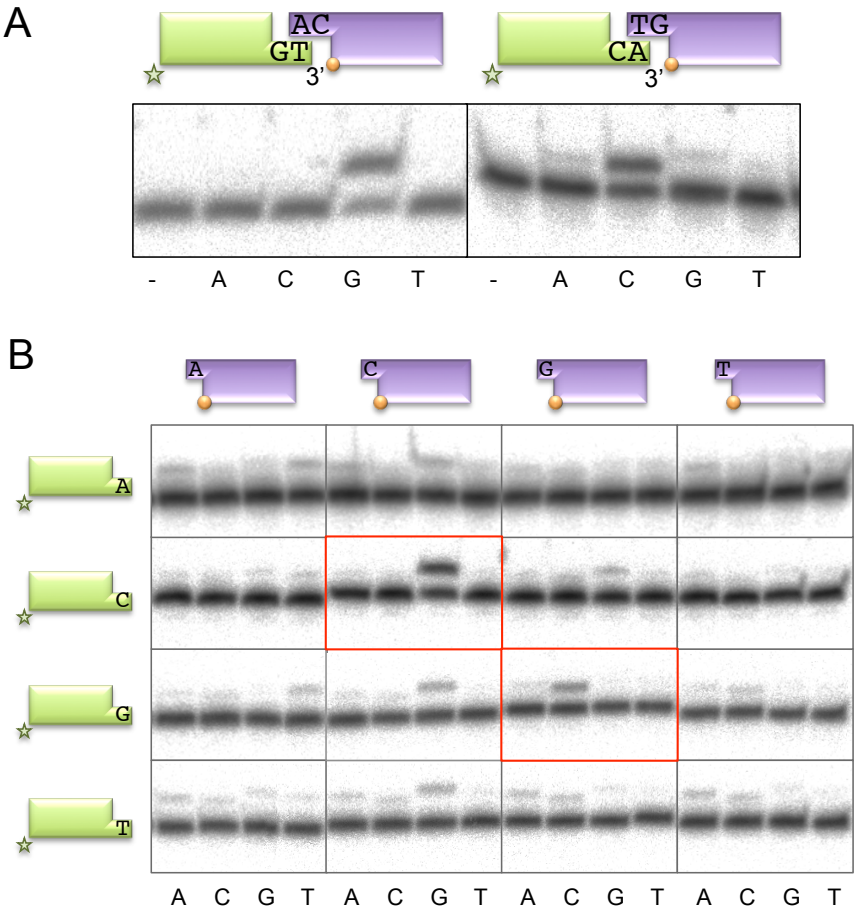



**Figure 26. Effect of the sequence context of the efficiency and fidelity of Pol*μ* during NHEJ.**
A) NHEJ assays of Pol*μ* (200 nM) performed as described in Materials and Methods, using two sets of compatible substrates: in the first one (left), the labeled substrate was formed by hybridization of 1TG and 1D-NHEJ (shown in green) and the cold substrate by hybridization of 1AC and 1D-NHEJ (shown in mauve). in the second one (right), the labeled substrate was formed by hybridization of 1AC and 1D-NHEJ (shown in green) and the cold substrate by hybridization of 1TG and 1D-NHEJ (shown in mauve). The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, dNTPs were added separately at 1 *μ*M in the presence of 2.5 mM $MgCl_2$. After electrophoresis, the labeled fragments were detected by autoradiography. B) NHEJ reactions performed as in A), with four sets of incompatible substrates in which the labeled one was formed by hybridization of either 1A, 1C, 1G or 1T with 1D-NHEJ (green) and the cold one, of either 2A, 2C, 2G or 2T with 2D-NHEJ (mauve). When indicated, dNTPs were added separately at 10 *μ*M in the presence of 2.5 mM $MgCl_2$.

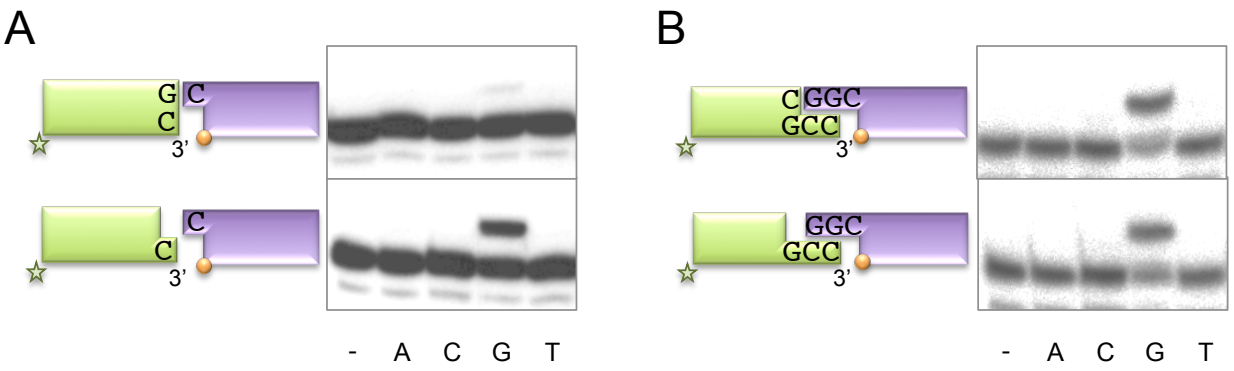



**Figure 27. Outcome after microhomology search: gaps or nicks.**
A) NHEJ assays of Pol$\mu$ (200 nM) performed as described in Materials and Methods, using two sets of incompatible substrates: in the first one (top), the labeled substrate was formed by hybridization of 1C and D-nick-G (shown in green) and the cold substrate by hybridization of 2C and 2D-NHEJ (shown in mauve). in the second one (bottom), the labeled substrate was formed by hybridization of 1C and 1D-NHEJ (shown in green) and the cold substrate by hybridization of 2C and 2D-NHEJ (shown in mauve). The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, dNTPs were added separately at 10 $\mu$M in the presence of 2.5 mM $MgCl_2$. After electrophoresis, the labeled fragments were detected by autoradiography. B) NHEJ reactions performed as in A), with two sets of compatible substrates: in the first one (top), the labeled substrate was formed by hybridization of 1-GCC and D-nick-C (shown in green) and the cold substrate by hybridization of 2-CGG and 2D-NHEJ (shown in mauve). in the second one (bottom), the labeled substrate was formed by hybridization of 1-GCC and 1D-NHEJ (shown in green) and

in every case tested, there was a basal reaction with dG (red asterisk), indicative of D3-D3 synapsis, which only provides dC as template; additionally, it can also be observed a similar level of insertion of a second nucleotide: the one directed by the corresponding D4 molecule (green asterisk; Fig. 25D, panels 1, 3&4). In the case of D4 X=C the unique insertion of dG is clearly a result of the sum of the contributions of both types of synapsis (Fig. 25D, panel 2). As it has been already shown, the presence of the 5'-P in the recessive strand of a DNA end targets this molecule to be bound by Pol$\mu$ as a template/downstream (T/D) substrate. Thus, to assign a T/D positioning only to the D4 molecule, while maintaining D3 positioned as primer, i.e to form the D3:D4 synapsis only, we restricted the presence of the 5'-P group to the D4 molecule, whereas the D3 substrate would present a hydroxyl group at this position. As shown in figure 25E, the insertion of dG contributed by the D3-D3 synapsis is now undetectable, while exclusive insertion of the nucleotide templated by the corresponding D4 molecule is efficiently achieved by Pol$\mu$ in each case. The same method was applied to try to channel the D4 molecule to the primer location, and allow D3 to act as template. When a 5'-P was present in the two molecules, exclusive insertion of ddG was detected on the D4 molecule, indicating that D4:D4 synapsis was preferentially formed (Fig. 25F). When the 5'-P group was restricted to the D3 molecule, and the D4 molecule contained a hydroxyl group, insertion of the D3-templated nucleotide was favored (Fig. 25G).

### 3. Sequence context effect on the efficiency and fidelity of Pol$\mu$-mediated end-joining

As substrates with a minimal complementarity (1 base pair connection) we used the same molecules as in figure 18, having two different 3' overhang sequences (being ...GT-3' and …CA-3'; see Fig. 26A). 5'-labeled molecules (in green), not harbouring a 5'-P, are favored to be selected as primers. Conversely, unlabeled T/D partner ends (in mauve) provide one connecting/complementary nucleotide and a free templating base proximal to a recessed 5'-P. In both examples, preferential extension of the labeled primer was directed "in trans" by the templating base neighbor to the 5'-P. However, the background insertion of the 3 wrong nucleotides significantly varied, being relatively high in the case of the ...CA-3' protruding primer (Fig. 26A). As substrates with null complementarity, we used the four different 1-nt protrusions to generate 12 situations of incompatible NHEJ out of the sixteen possible combinations (Fig. 26B). As a result, we concluded that there is a gradation of both efficiency and accuracy when Pol$\mu$ is confronted with different incompatible DNA ends: some ends allow a very efficient and error-free reaction (G and C as template or primer), while the levels of

polymerization and fidelity achieved on others (T and A as template or primer) are very low.

## 4. Structure of the DNA substrates

### 4.1 Space matters: nicks versus gaps

When two DNA ends are brought together, and if the 3' overhangs are different in length, it is likely that instead of the formation of two adjacent gaps, a nick could be the only (and minimal) discontinuity occurring in the template strand, opposite to the gap that needs to be filled by the polymerase. If a ligase is able to act first, sealing the nick, the remaining repair would be a short gap filling. Alternatively, a PolX could be more efficient to bridge such DNA ends, preceding any ligase action. Thus, we wanted to test if Pol*μ* was able to deal with this situation, and for that we used two different blunt-ended substrates as primer and two different DNA cold molecules whose 1 nt-protrusion (either a dG or a dC) close to a recessive 5'-P, can be used as template (Fig. 27A). When comparing the activity of Pol*μ* on these "nick/gap" *versus* "gap/gap" substrates, the observed activity is around 10-fold worse in the first set with respect to the second set of substrates. We can conclude that the polymerase takes advantage of the presence of a second gap in the template strand, that could provide an effective interacting site, perhaps *via* Loop 1, to correctly position the two incompatible ends (ref Nick McElhinny and Juarez). To further test this idea, we used complementary DNA end substrates that upon connection form a nick in the template strand, but -3 base pairs away from the polymerization site (Fig. 27B). In this situation, in which Loop 1 would not be playing such a direct role (see below), the two primers appear to be similarly extended, irrespective of the presence of either a nick or a gap in the template strand.

### 4.2 Substrates with 2 gaps: separation and length of the gaps.

Based on the same premise, i.e. 3'-protrusions having different length, formation of two gaps separated by a distance longer than one base pair could occur; so, we decided to test the efficiency of Pol*μ* on NHEJ substrates that, upon connection, can form two gaps separated by none, 1, 2 and 3 complementary base pairs. For these assays we used two series of oligonucleotides, consisting of a common downstream oligo, and different templates which provide 3' overhangs of increasing lengths: from 1 to 4 nucleotides. The sequences of the 3'-protrusions used are: …G, …GT, …GTT and … GTTT for the first series, and …C, …CA, …CAA and …CAAA for the second series. Thus the two series can be combined together to achieve different levels of microhomology and distance between the gaps

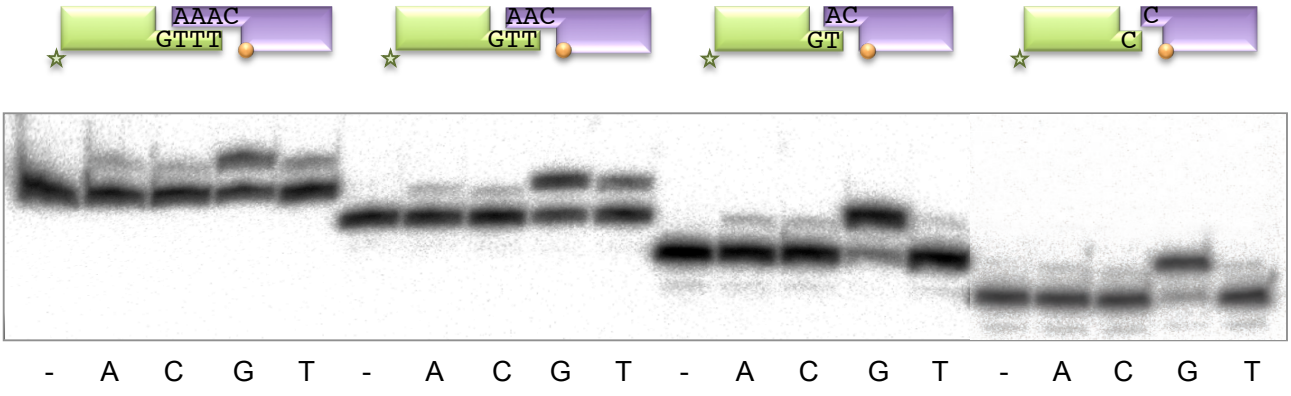



**Figure 28. Distance between the two gaps and effect on efficiency of the NHEJ reaction.**
NHEJ assays of Pol*μ* (200 nM) performed as described in Materials and Methods, using four sets of substrates: in the first one (left), the labeled substrate was formed by hybridization of 1-TTTG and 1D-NHEJ (shown in green) and the cold substrate by hybridization of 1-AAAC and 1D-NHEJ (shown in mauve). in the second one, the labeled substrate was formed by hybridization of 1-TTG and 1D-NHEJ (shown in green) and the cold, of 1-AAC and 1D-NHEJ (shown in mauve). in the third one, the labeled substrate was formed by hybridization of 1-TG and 1D-NHEJ (shown in green) and the cold substrate by hybridization of 1-AC and 1D-NHEJ (shown in mauve). in the fourth one, both the labeled and the cold substrates were formed by hybridization of 1-C and 1D-NHEJ (shown in green). The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, dNTPs were added separately at 10 *μ*M in the presence of 2.5 mM $MgCl_2$. After electrophoresis, the labeled fragments were detected by autoradiography.

(0, 1, 2 or 3 base pairs, Fig. 28). As shown in figure 28, Pol*μ* preferred the substrate containing a 3'-protrusion of 2 nt, that provides one base pair of microhomology once joined (...GT *versus* …CA). Overhangs longer than this are less efficiently joined, as evidenced by the reduced activity of Pol*μ* on these substrates. This is a paradox assuming that the higher the homology, the easier the reaction. For Pol*μ* this assumption seems to be wrong, as our observations indicate that the closer the distance between the two gaps, the higher the activity achieved by the polymerase. This effect is attributable to Loop 1, a protein domain that is believed to keep the connection between the DNA ends, perhaps by stabilizing the second gap when the two ends are joined together. In that case, a Loop 1-mediated stabilization would be only possible when the location of the second gap is at an optimal distance, close to the position of Loop 1. When the distance between the gaps is excessive, and since the polymerase remains attached to the downstream side of the first gap by its 5'-phosphate, the second gap would be located too far away from Loop 1, and thus this domain would not be able to stabilize end-joining. So, it appears that Loop 1 is specially suited to help in the stabilization of very short protrusions, with little or null complementarity.

**4.3 2-nt gaps: dislocation in NHEJ context**

Pol*μ* exhibits a strong mutator phenotype, showing a very high misincorporation rate, and is thus one of the most unfaithful polymerases known in higher eukaryotes (Dominguez et al., 2000). The strong mutator activity of Pol*μ* when copying a DNA template resides on its ability to create or accept distortions and realignments of both primer and template strands, being largely dependent on sequence context (Ruiz et al., 2004; Zhang et al., 2001). It was found that many of the alleged misincorporations by Pol*μ* were the result of base-pairing between the incoming nucleotide and complementary positions close to the templating base. Analysis of different sequence contexts showed that the most extreme behavior corresponded to a "slippage" mechanism, described initially for Polß on a template/primer (Kunkel, 1985). This slippage mechanism implies the formation of a template dislocation, maintained by the new pairing stablished by the 3' terminus of the primer to the n+1 position, given that n and n+1 bases in the template strand are identical (Fig. 29A). Pol*μ*'s propensity to dislocate the template strand is such that it preferentially inserts the nucleotide complementary to the n+2 position of the template, both in an open template/primer and in a 2nt-gap (Juarez, 2006). After the first incorporation, the extended primer could be realigned with the n+1 template (likely forming a mismatch), that could be eventually elongated to fulfill the gap, in a reaction directed again by the n+2 templating base. Both the insertion after dislocation of the template and direct insertion without repositioning of the template strand are possible on this type of substrate. However, it is not strictly necessary that the bases n and n+1 of the template are identical for a template dislocation to occur. Thus, there is a dislocation model in which the stabilization of the misaligned intermediate is mediated by stacking interactions with the incoming dNTP (Fig. 12A; Kunkel, 1985), and/or by the intrinsic formation of a polymerase pre-ternary complex (Brissett et al*.,* 2011; chapter X, this PhD work). This mechanism ("dNTP selection-mediated") is not as efficient as the "slippage-mediated" alternative, but is considerably improved when the substrate is a small gap. This observation emphasizes the importance of a stable binary complex for this reaction to occur. We wanted to check whether this dislocation capacity, that is exhibited by Pol*μ* in the context of perfectly matched DNA gaps, could also be used to solve certain special situations, as would be case of NHEJ reactions that include the formation of 2 nt gaps. We have also shown that the sequence context plays a very important role on the efficiency and accuracy of the NHEJ reactions, and therefore, we decided to test if dislocation of the template, typical of Pol*μ*, could be operating even in the case of an unfavorable sequence context.

For this we used a variation on the D3 and D4 NHEJ substrates used before: in this case the length of the 5-P containing gap is 2 nt and 1 nt for the second gap (Fig. 29B). The sequence of the two templating bases of the first gap always includes a C, which directs nucleotide incorporation in an efficient and accurate fashion ("strong" templating base). In a first set of substrates, the C is positioned at the n+1 position of the template, and the n+2 base is changed to A, C, G or T (X). In a second set of substrates, the C is located at the n+2 position of the

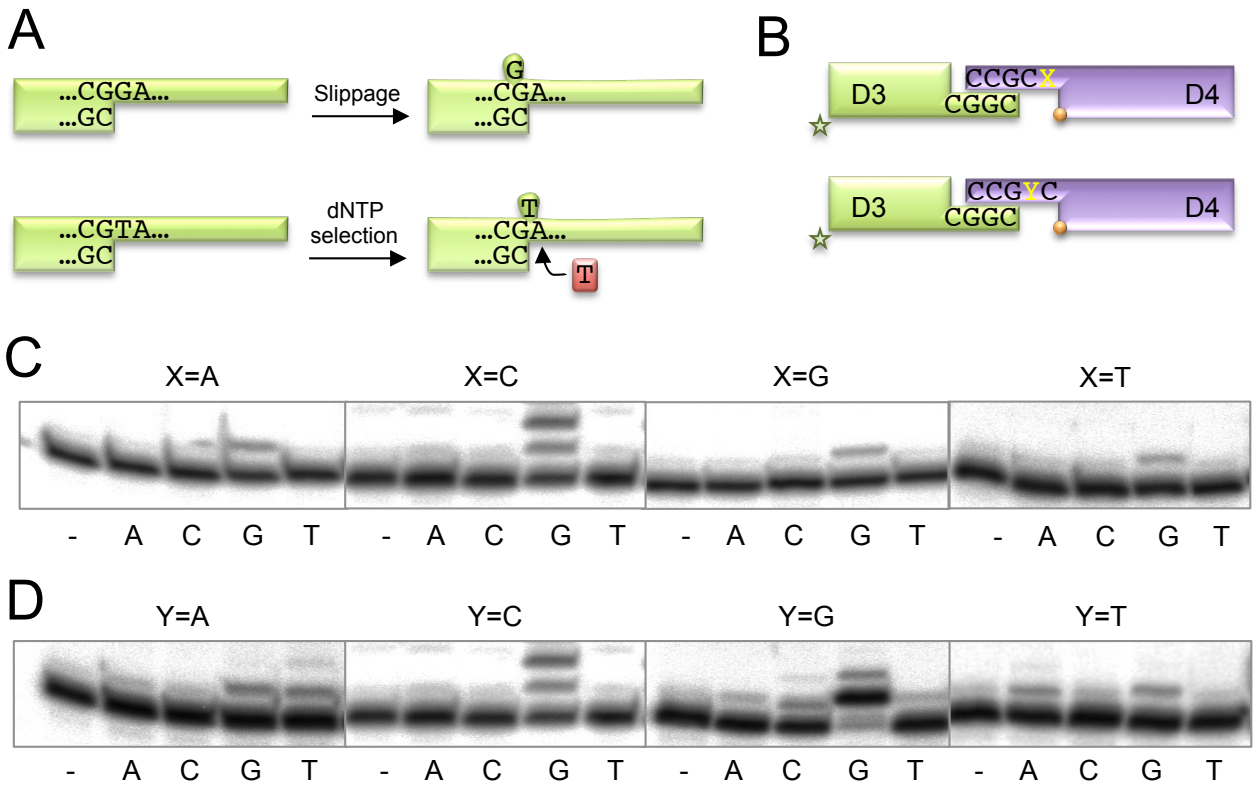



**Figure 29. Frameshifts during Pol$\mu$-mediated NHEJ.**
A) Schemes of the two mechanisms by which Pol$\mu$ can produce frameshifts. See main text for details. B) Scheme of the substrates used in the assays in C)&D). The X and Y positions will be changed to any of the four possibilities. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. C&D) NHEJ reactions performed as described in Materials and Methods, with 200 nM Pol$\mu$ and a set substrates indicated in B). When indicated, dNTPs were added separately at 10 $\mu$M in the presence of 2.5 mM $MgCl_2$. After electrophoresis, the labeled fragments were detected by autoradiography.

template, whereas n+1 base is changed to A, C, G or T (Y). The results showed that Po$\mu$ dislocates the template strand, to insert the incoming nucleotide opposite to the n+2 position, only when the sequence context favors this dislocation, i.e. when the C is in this n+2 position of the template, and the n+1 position is occupied by a "weak" templating base such as A or T, except in the case of Y=G, in which the dislocation is slippage-mediated, and strongly favored (Fig. 29C). If the "strong" templating base is placed at the n+1 position of the template, the levels of dislocation are greatly diminished (Fig. 29D).

**5. Non-aligned ends: dealing with mismatches, bubbles, flaps.**

Another physiological situation that could occur when two DNA ends are brought together, is the formation of a variety of distortions in both strands at the connexion. This distortions could include mismatched base pairs, the formation of 3'-flaps of ssDNA, the formation of "bubbles" of extrahelical nucleotides, etc (Brissett et al., 2007). We wanted to asses NHEJ efficiency when these imperfections are present at Pol$\mu$-mediated synapsis. We used the same NHEJ substrates used in the experiment shown in figure 8E, bearing three base pairs of complementarity, but with variable protrusions to include a wide variety of distortions in the synapsis. The results showed that Pol$\mu$ can deal with complementary ends containing 1nt "bubbles" at the n-1 and n-2 positions of the primer with the same efficiency as with perfectly complementary ends (Fig. 30A). We confirmed this result by having all 4 options as template to make sure that the synapsis was being correctly formed (Fig. 30B). The presence of a mismatch in the connexion does not seem to have any deleterious effect either (Fig. 30C, second panel). However, the presence of a 1nt 3'-flap, or a mismatched 3' primer terminus, totally inhibits polymerization on that DNA end, but not on the other (Fig. 30C, third and fourth panels).

**6. Damaged ends: 8oxoG in DNA ends.**

One of the most common oxidative damages in the cell is the modification of the nitrogenated base

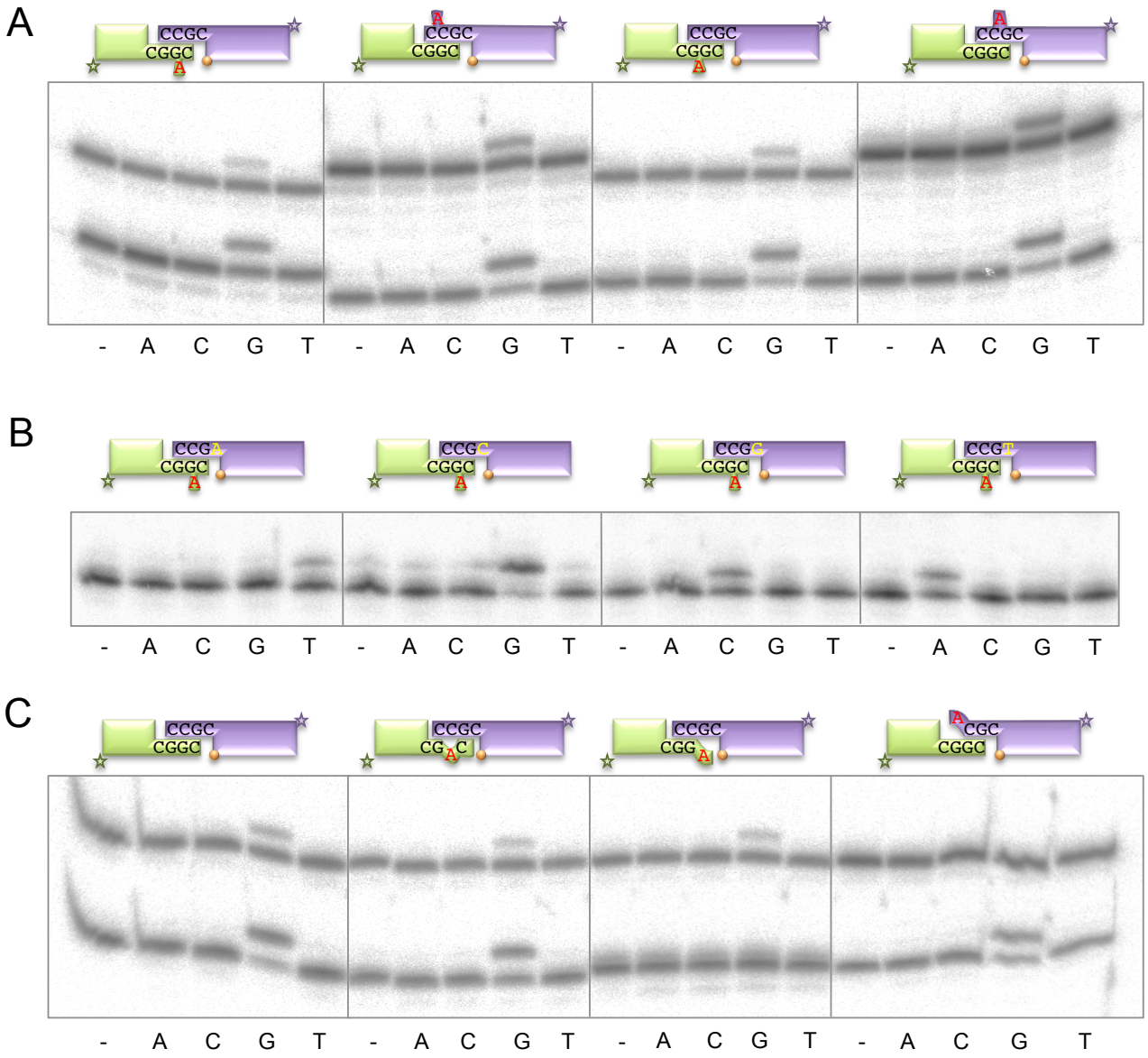



**Figure 30. Pol$\mu$-mediated NHEJ: non-aligned ends.**
A) NHEJ reactions performed as described in Materials and Methods, with 200 nM Pol$\mu$ and using labeled compatible substrates: the short substrates were formed by hybridization of either D3-C (second and fourth schemes), D3-BB1 (first scheme) or D3-BB2 (third scheme) with D1 (shown in green) and the long substrates, by hybridization of either D4-C (first and third schemes), D4-BB1 (second scheme) or D4-BB2 (fourtth scheme) with D2 (shown in mauve). The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, dNTPs were added separately at 1 $\mu$M in the presence of 2.5 mM $MgCl_2$. After electrophoresis, the labeled fragments were detected by autoradiography. B) NHEJ reaction performed as A), with the first set of substrates, and chamging the sequence of the templating base in the long (cold) substrate as indicated. C) NHEJ reactions performed as in A), using labeled compatible substrates: the short substrates were formed by hybridization of either D3-C (second and fourth schemes), D3-MM (first scheme) or D3-FLAP (third scheme) with D1 (shown in green) and the long substrates, by hybridization of either D4-C (first to third schemes) or D4-FLAP (fourth scheme) or D4-BB2 (fourtth scheme) with D2 (shown in mauve). The orange

guanine to 7,8-dihydro-8-oxoguanine (8oxoG), which happens with a frequency of $10^3$-$10^4$ times per cell and day (Beckman and Ames, 1997; Helbock et al., 1998). This lesion is very dangerous to the cell: it is highly mutagenic, producing transitions G:C ➔ T:A (Moriya, 1993). This is due to its ability to pair with cytosine but also with adenine (Fig. 14A). In the 8oxoG:C base pair, the glycosidic bonds of both bases are in an *anti* conformation and stabilization of the pairing relies on three hydrogen bonds (Lipscomb et al., 1995; Oda et al., 1991). On the other hand, when 8oxoG pairs with A, the glycosidic bond of the damaged base is in *syn* conformation and forms a Hoogsteen pairing, stabilized by two hydrogen bonds (Kouchakdjian et al., 1991; McAuley-Hecht et al., 1994). Another detail that helps to explain the mutagenicity of 8oxoG is that, when present in the DNA, either paired with C or

with A, the B configuration of the double helix is maintained (Kouchakdjian et al., 1991; Lipscomb et al., 1995; McAuley-Hecht et al., 1994; Oda et al., 1991), although in the case of the pair 8oxoG·dCMP the DNA carbon backbone shows certain perturbations (Lipscomb et al., 1995).

During translesion synthesis there are two relevant steps: firstly, the insertion opposite the damaged base has to occur, and secondly, the pair generated has to be extended. The ability to perform one or another (or both) of these two steps by a polymerase is defined by two parameters: the efficiency and the fidelity of nucleotide incorporation when the polymerase is either copying or extending the damage. When copying a damage, the relative efficiency of nucleotide insertion is calculated in comparison to that observed when copying an undamaged template. Fidelity of translesion is defined by the ability to insert the correct nucleotide in front of the damaged templating base (i.e. dC in front of 8oxoG; in that case, insertion of dA is particularly relevant). When extending base pairs containing a damaged base, relative efficiency is determined in comparison with the extension of the corresponding undamaged base pairs, and fidelity refers to the increased ability to extend the correct pair (8oxoG:C) compared to the incorrect pair (8oxoG:A).

In a conventional template/primer substrate, the three template-dependent members of the human X family of polymerases preferentially insert the correct nucleotide, complementary to the first templating base closest to the 3'-terminus of the primer. However, Pol*μ* also inserts efficiently the nucleotide complementary to other templating bases available, inducing a dislocation of the template (Ruiz et al., 2004; Zhang et al., 2001). When testing their behaviour in a template/primer substrate in which the first templating base is 8oxoG profound differences are observed: Polß preferentially incorporates the "correct" nucleotide, dC, although less efficiently than in front of an undamaged dG template (Miller et al., 2000); Polλ incorporates not only dC, but also dA, opposite 8oxoG in a very

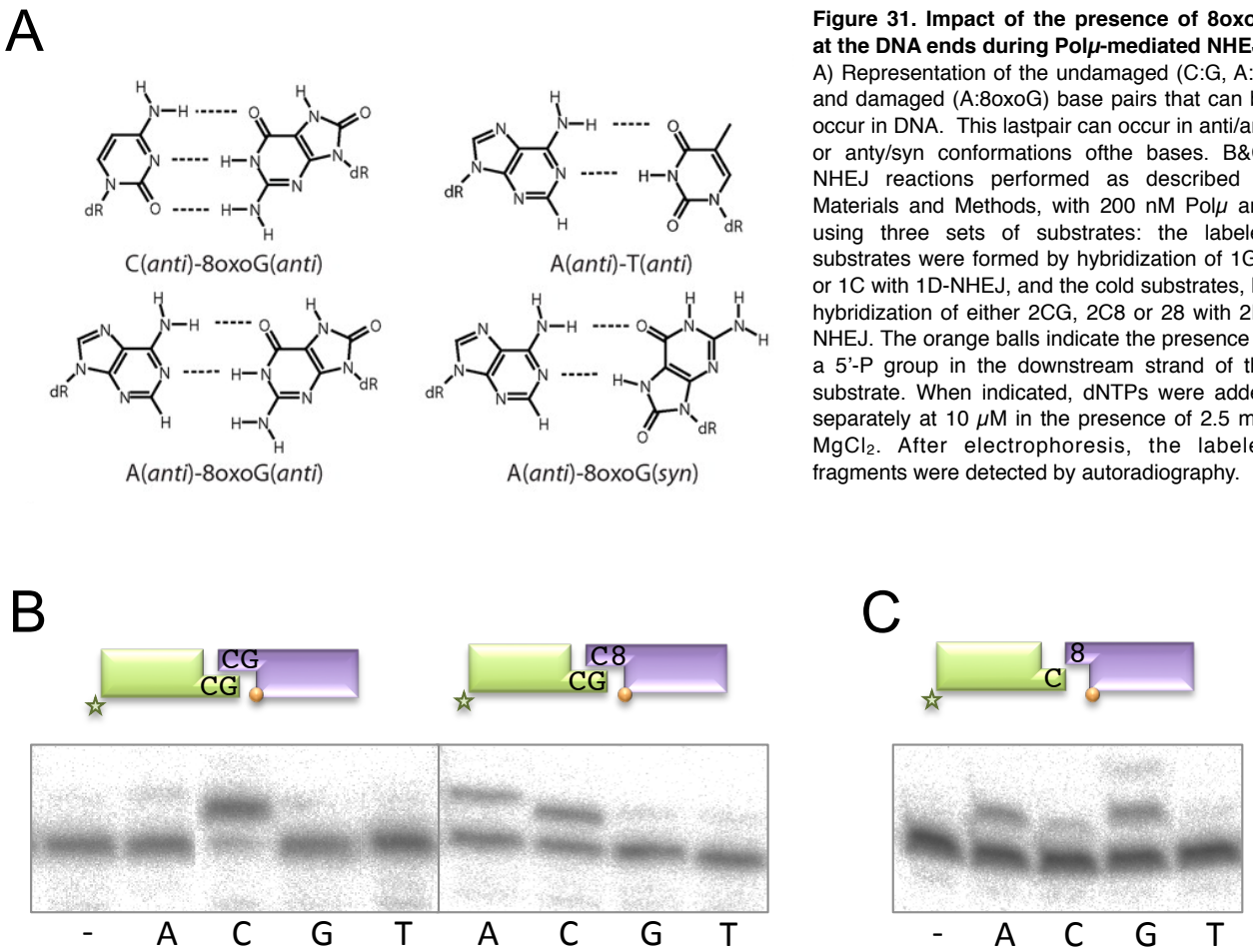



**Figure 31. Impact of the presence of 8oxoG at the DNA ends during Pol*μ*-mediated NHEJ.** A) Representation of the undamaged (C:G, A:T) and damaged (A:8oxoG) base pairs that can be occur in DNA. This lastpair can occur in anti/anti or anty/syn conformations ofthe bases. B&C) NHEJ reactions performed as described in Materials and Methods, with 200 nM Pol*μ* and using three sets of substrates: the labeled substrates were formed by hybridization of 1GC or 1C with 1D-NHEJ, and the cold substrates, by hybridization of either 2CG, 2C8 or 28 with 2D-NHEJ. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, dNTPs were added separately at 10 *μ*M in the presence of 2.5 mM $MgCl_2$. After electrophoresis, the labeled fragments were detected by autoradiography.

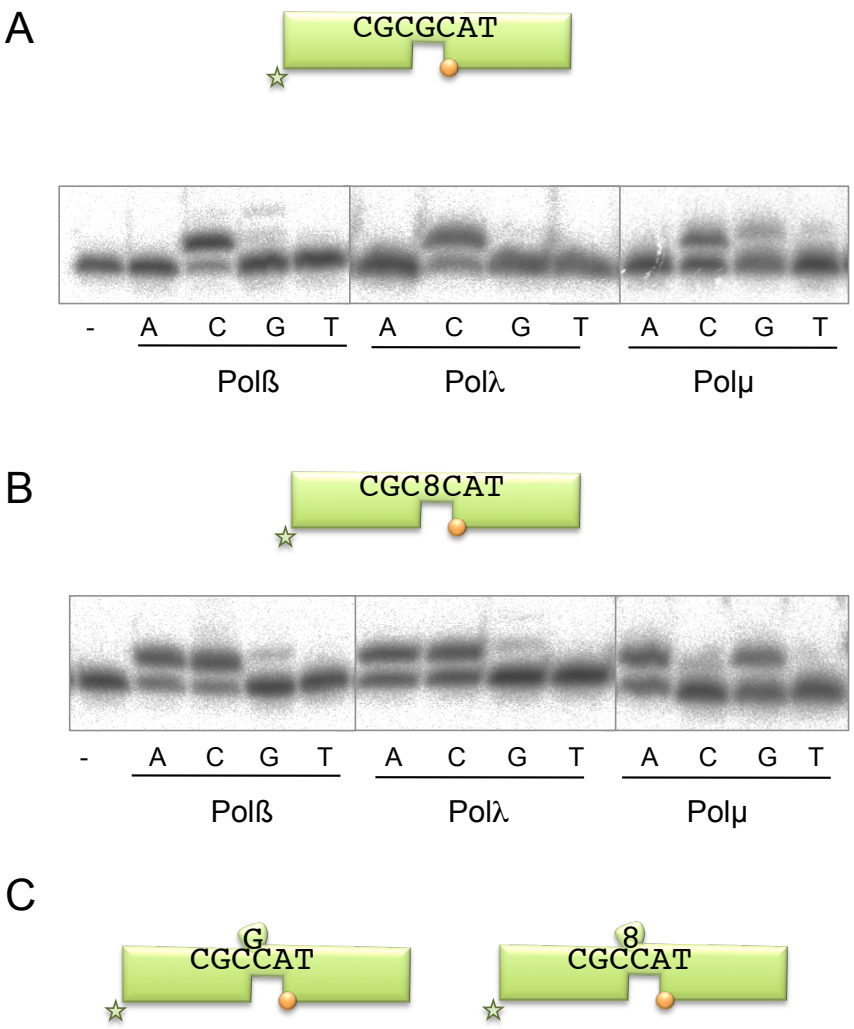



**Figure 32. 8oxoG as template during gap-filling by X family polymerases.** A) Gap-filling assay of Polß, Polλ or Pol*μ* (25 nM) with an undamaged template formed by hybridization of the oligos pBER, dBER and T4, in the pressence of 10 nM of each of the four dNTPs and 2.5 mM $MgCl_2$. After electrophoresis, the labeled fragments were detected by autoradiography. B) Gap-filling assay of Polß, Polλ or Pol*μ* (25 nM) with a damaged template contaiing an 8oxoG as the templating base, formed by hybridization of the oligos pBER, dBER and T8, in the presence of 10 nM of each of the four dNTPs and 2.5 mM $MgCl_2$. After electrophoresis, the labeled fragments were detected by autoradiography. C) Schemes showing the possible arrangements of the substrates in A) and B) induced by Pol*μ* that would result in a -1 frameshift.

efficient reaction, reaching levels similar to the incorporation of dC in front of a normal dG (Picher and Blanco, 2007); Pol*μ*, on the other hand, inserts only dA in front of 8oxoG in this kind of substrates, although the dislocation mechanism is enhanced by this damaged templating base and, consequently, the level of incorporation of the next nucleotide is increased with respect to that obtained when the first templating base is a normal dG (Picher and Blanco, 2007). In a more complex but physiological scenario, modifications as 8oxoG could be eventually part of the synapsis required to bridge DNA ends, and therefore, the capacity of Pol*μ* either to insert nucleotides opposite 8oxoG or to extend base pairs containing such a damage could be beneficial for its role in NHEJ.

**6.1 8oxoG as templating base: efficiency and fidelity of incorporation during the first step of NHEJ.**

In the first part of this study we wanted to focus on the behaviour of Pol*μ* when using the damaged base 8oxoG as part of the connection of the two ends during the first gap-filling of the NHEJ reaction. We used a set of NHEJ substrates similar to those used above, with modified 3'-overhangs. Two situations have been studied: in the first and "easy" case one base pair of microhomology is formed upon synapsis; in the second and more "difficult" case no complementarity exists among the two ends. During synapsis, either an undamaged dG or the damaged 8oxoG is at the templating position, closest to the 5'-P of the downstream strand (present only in this 8oxoG-containing substrate to force its role a template). Figure 32B&C shows the results obtained with these substrates. In the case of the complementary ends, Pol*μ* efficiently and faithfully incorporates dC in front of dG, but when 8oxoG acts as the templating base during this NHEJ reaction, in addition of the predicted misincorporation of dA, the correct nucleotide (dC) was unexpectedly inserted at about the same rate. In the case of non-complementary ends this preference changes, since in every case tested Pol*μ* preferred to use the

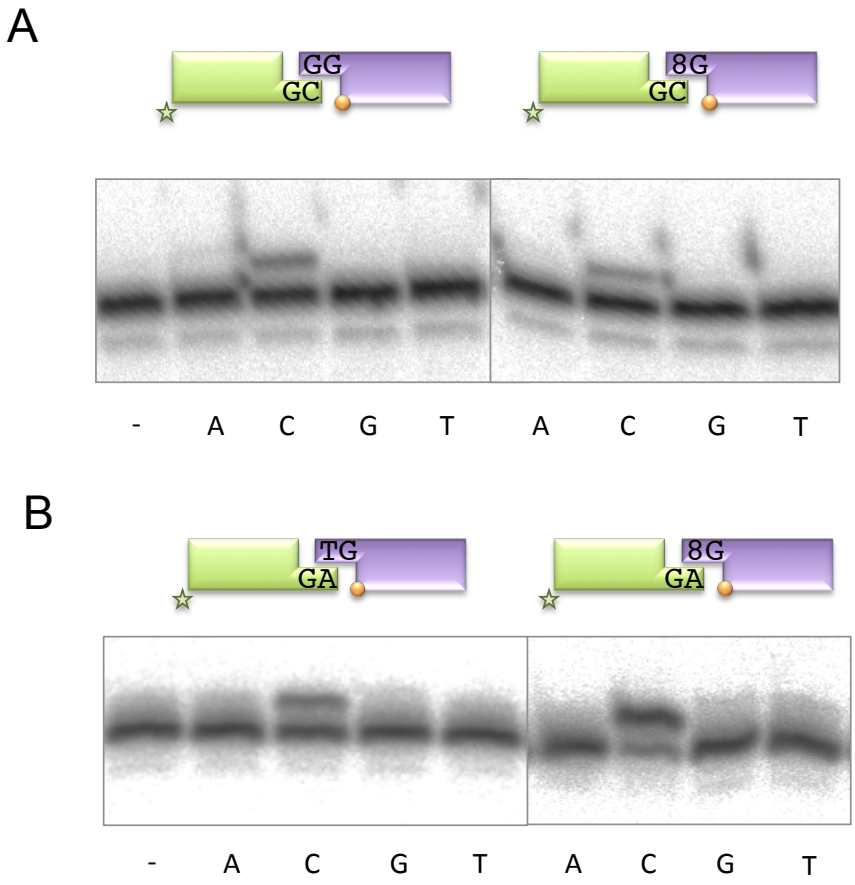



**Figure 33. 8oxoG at the connection of two DNA ends during Polμ-mediated NHEJ.**
A&B) NHEJ reactions performed as described in Materials and Methods, with 200 nM Polμ and using three sets of substrates: the labeled substrates were formed by hybridization of 1CG or 1AG with 1D-NHEJ, and the cold substrates, by hybridization of either 2GG, 2G8 or 2TG with 2D-NHEJ. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, dNTPs were added separately at 10 μM in the presence of 2.5 mM $MgCl_2$. After electrophoresis, the labeled fragments were detected by autoradiography.

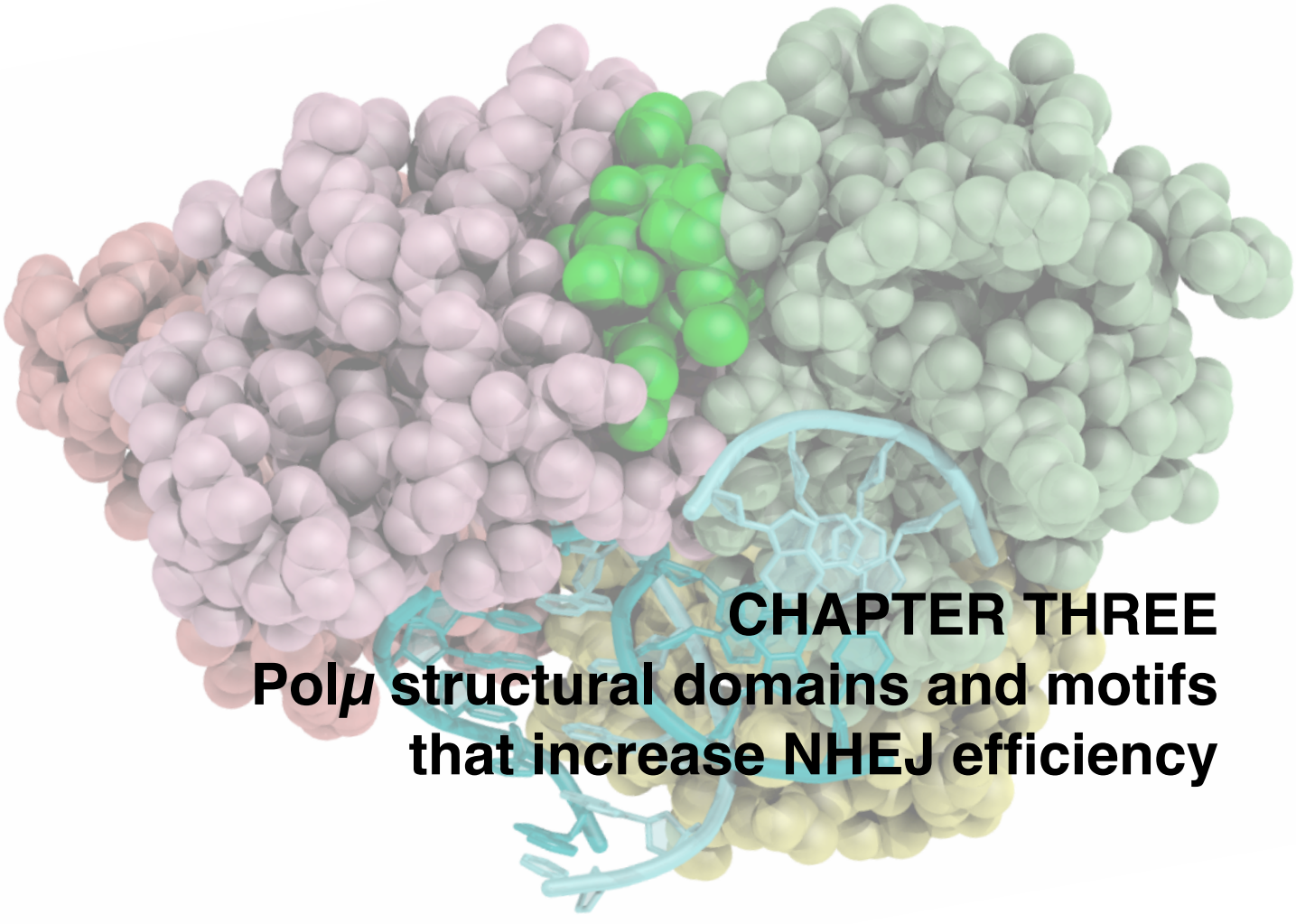
undamaged end both as primer and as template (despite it is not bearing a 5'-P in the downstream strand) than using the damaged end (that contains a 5'-P) as template. As an example, in the case shown in figure 32C, we could observe incorporation of ddG (likely a templated event resulting from NHEJ of two undamaged ends) (ddG), but importantly, we were also able to detect the incorporation of nucleotides in front of 8oxoG (ddA and ddC).
From the data obtained we learned that in both complementary and non-complementary ends Polμ is able to tolerate 8oxoG during NHEJ. In both cases, error prone insertion of dA, and even error-free insertion of dC, occurs in front of the damaged base, as efficiently as during similar end-joining events with undamaged templating bases.

### 6.2 8oxoG as templating base: efficiency and fidelity of insertion during the second step of NHEJ.

We next tested 8oxoG as a template in a DNA gap. In this case, the three polymerases Polß, Polλ and Polμ were tested in parallel since this substrate is representative of the second step of the NHEJ reaction, once the first gap has been filled and ligated, but is also a substrate of the BER pathway that involves Polß and Pol. Results showed that the three polymerases are able to incorporate the correct nucleotide (dCTP) in front of an undamaged base (dG; Fig. 32A), although Polμ also incorporates the incorrect nucleotide dGTP, probably due to its ability of realigning the template strand, causing in this case a +1 frameshift. In the case of a damaged templating base (8oxoG) the polymerases displayed a very different behaviour: they all readily incorporate dATP in front of the damaged template, but Polß and Polλ also incorporated the correct dCTP, at the same rate as dATP, as it has already been shown (Picher and Blanco, 2007; Fig. 32B). On the other hand, Polμ does not incorporate dC, but dG, again because of the primer realignment (Fig. 32B, right panel, and Fig. 32C). This behaviour of Polμ is highly mutagenic, as it never incorporates the correct nucleotide (dC) in front of 8oxoG. This is in clear contrast to its signature of nucleotide incorporation during the first step of NHEJ, when the correct dC is efficiently incorporated.

### 6.3 8oxoG in the connection: efficiency and fidelity of the extension of matched (8:C) and mismatched (8:A) pairs during the first step of NHEJ.

On regular template/primer substrates, the three template-dependent DNA polymerases of family X (Poß, Polλ and Pol*μ*) perform effectively the extension of a dG·dCMP pair, with Polλ being the most efficient due to its lower $K_m$ per nucleotide (Picher and Blanco, 2007). However, the behaviour of each DNA polymerase is different when assessing extension of the possible pairs generated during the insertion step in front of a modified base 8oxoG (8oxoG·dCMP and 8oxoG·dAMP). Polß and Polλ have a high fidelity of extension, as they mainly prefer to extend the correct pair (8oxoG·dCMP), while Pol*μ* has the lowest level of fidelity, and similarly extends the correct and the mutagenic pairs (Picher and Blanco, 2007). We wondered if this Pol*μ* behaviour would be maintained when the extension of the damaged pairs has to occur in the context of a NHEJ reaction. When extending a 8oxoG·dCMP pair occurring at the connection of two DNA ends, Pol*μ* showed the same efficiency as with the undamaged dG·dCMP pair (Fig. 33A). Strikingly, when extending the mutagenic pair 8oxoG·dAMP (Fig. 16B), occurring at the connection of two DNA ends, Pol*μ* was more efficient than when extending both the undamaged dT·dAMP and dG:dCMP pairs, and the damaged 8oxoG·dCMP pair. These observations finally lead to the conclusion that Pol*μ* has a highly mutagenic behaviour in the context of NHEJ of oxidative damaged ends, since it incorporates preferentially dAMP in front of 8oxoG at both the first and second stages of a NHEJ reaction, and it also extends preferentially the mutagenic pair 8oxoG·dAMP.



# CHAPTER THREE
# Pol*μ* structural domains and motifs that increase NHEJ efficiency

**1. YACQR motif: A closed conformation of the polymerase core is optimal for NHEJ**

Structural characterization of Polß and several DNA polymerases from families A and B suggests that dNTP binding induces large conformational changes affecting the relative positioning of both fingers and thumb subdomains (Doublie et al., 1999; Li et al., 1998; Sawaya et al., 1997). For some polymerases, this change from an "open" to a "closed" conformation and other minor conformational changes might serve as multiple fidelity checkpoints (Joyce, 2004) remodeling the active site and the nascent base pair binding pocket to trigger optimal catalysis. Unlike Polß, that undergoes huge domain motions implicating the 8 kDa and thumb subdomains (Fig. 34A), structural characterization of Polλ indicates that its catalytic cycle does not involve major subdomain motions (Fig. 34B), i.e., Polλ is in a closed conformation prior to dNTP binding (Garcia-Diaz et al., 2005). Informative crystal structures, leading to this conclusion in the case of Polλ (i. e. binary with gapped DNA *versus* pre-catalytic ternary with gapped DNA and incoming nucleotide; see figure 34B), do not exist for Pol*μ*; For the latter, only a ternary complex with gapped DNA and incoming nucleotide has been solved (Moon et al., 2007; PDBid: 2IHM; Fig 34B). Despite the lack of enough structural information, comparison of the existing structures and multiple amino acid sequence alignments guided us to identify a motif that is present in the catalytic core of Pol*μ*, TdT and Polλ, but absent in Polß, that could be in charge of maintaining this closed conformation of the polymerase core throughout the catalytic cycle. This motif, formed by the five amino-acids Tyr-Ala-Cys-Gln-Arg (YACQR) in Pol*μ*s and TdTs, is extraordinarily conserved among different species (Fig. 34C). In Polλs, the conserved motif is slightly different, but includes again five amino-acids: Trp-Val-Cys-Ala-Gln (WVCAQ). This motif, located at the very N-terminal part of the core, was not included in the Polλ core that has been crystallized (Garcia-Diaz et al., 2004; PDBid:1XSL). Fortunately, the murine Pol*μ* core construct whose structure has been solved does include the YACQR motif, and therefore, we could analyze the interaction of these amino-acids, located at the 8 kDa subdomain, with residues located in the thumb subdomain (Fig. 34D). Tyr$^{141}$ of the motif is interacting with thumb residues Gln$^{452}$ and Glu$^{453}$ through their side chain, as well as with Gln$^{144}$ of the YACQR motif, through the backbone of both residues. Arg$^{145}$ of the motif is interacting with other residues located at the 8 kDa subdomain (Pro$^{139}$, Ala$^{140}$, Ser$^{184}$ and Ser$^{188}$), thus helping to maintain the general architecture of the motif. We speculate that these interactions would be involved in the maintenance of the closed conformation of the core, and consequently in the correct positioning of the DNA, specifically during the first step of a NHEJ reaction in which the protein has to interact with two different DNA substrates. A constitutively closed core could be critical to interact with the downstream DNA end, that would be bound by its 5'-phosphate group to the 8 kDa domain of the polymerase. So, if the protein architecture is not properly maintained, the DNA ends would no be correctly located for the NHEJ reaction. To test this idea, we mutated each of the three most conserved residues of the YACQR motif of human Pol*μ*. The point mutations were designed as follow: Tyr$^{141}$ was mutated either to phenylalanine (Y141F) to maintain the overall shape of the aminoacid but to lose the reactive group, or serine (Y141S) to get the opposite result; Cys$^{143}$ was mutated to glycine (C143G) in order to abolish the interactions stablished by this residue; and Arg$^{145}$ was mutated to either lysine (R145K), to maintain the positive charge, or alanine (R145A), to eliminate any function of this residue. We also targeted another residue, Asn$^{153}$, that might be involved in this conformational function, complementing the role of the YACQR motif. This residue is located also in the 8 kDa subdomain, and although it is not establishing direct interactions with any residue located in the thumb domain, it is part of a network of interactions that maintain the general conformation of this part of the 8 kDa subdomain (Fig. 34D). This structural function seems to be crucial given the invariant conservation of this residue among Pol*μ*s, TdTs and Polλs of different species (Fig. 34C). So, we decided to make a point mutation on Asn$^{153}$ of human Pol*μ* changing it to glycine (N153G), to test the hypothesis that this destabilization could possibly affect the function of the YACQR motif. Other Pol*μ* residue possibly connected to this structural function would be Arg$^{448}$, which is located in the thumb subdomain and interacts with residues Glu$^{445}$, also in the thumb, and Tyr$^{204}$, located in the 8 kDa domain. The effects of mutating this residue on Pol*μ* activity will be

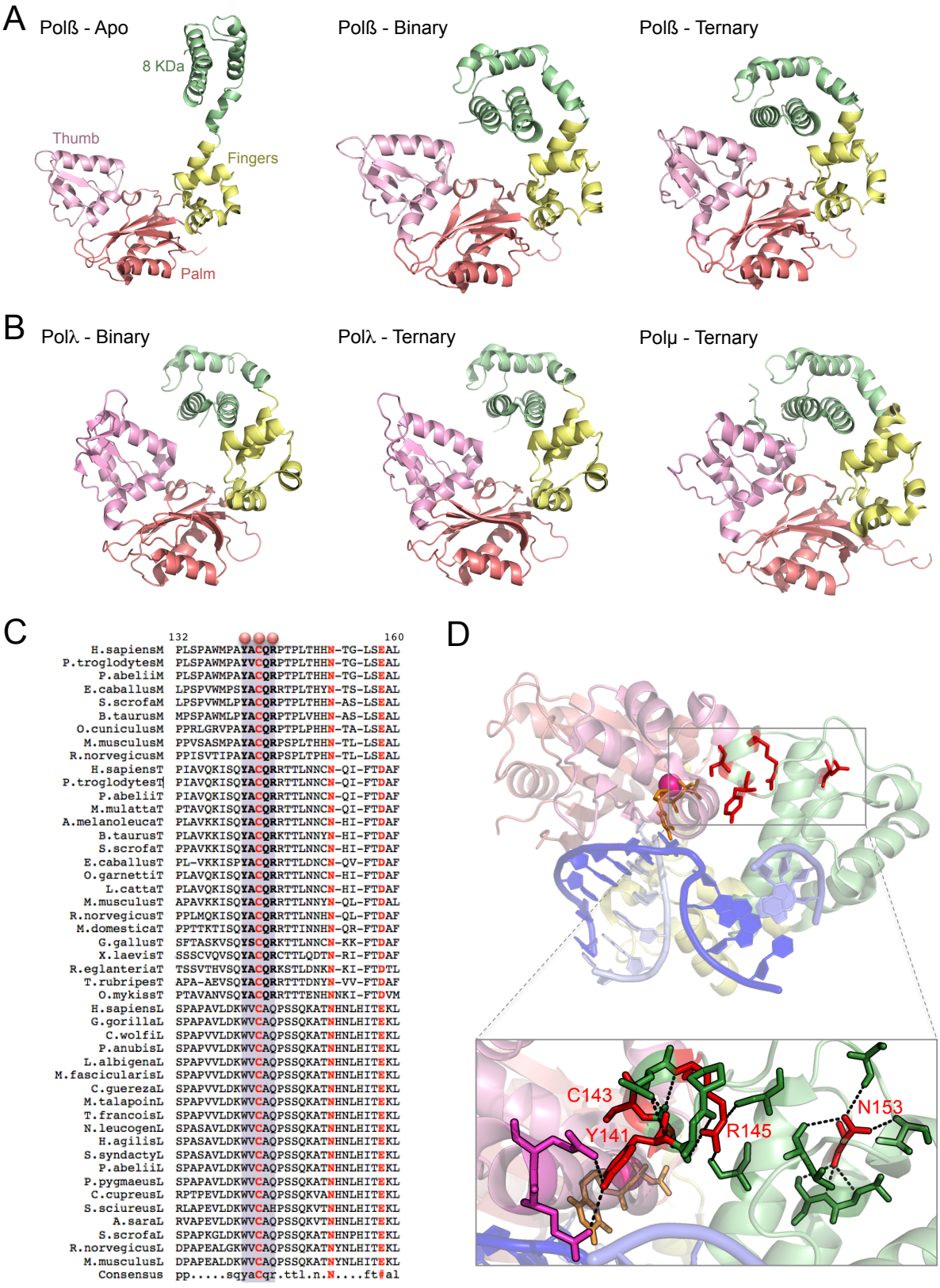
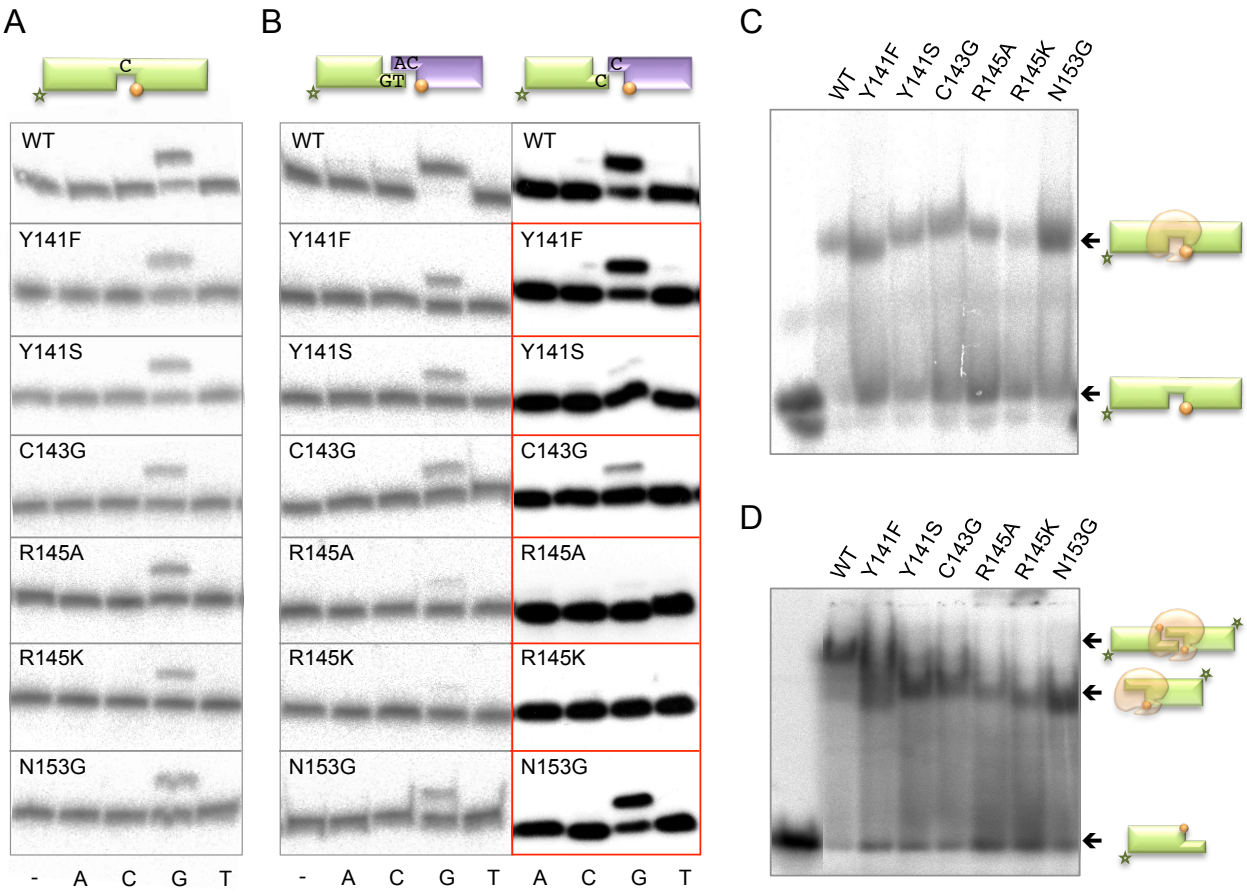
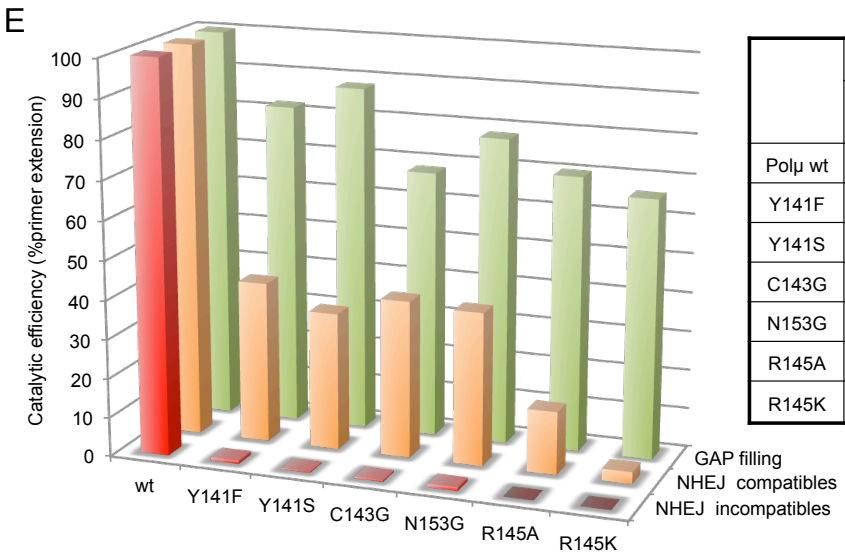



**Figure 34. The "brooch" motif (YACQR): level of conservation and implications for the closing of the polymerase core.** A) Cartoon representations of the apoenzyme, binary and ternary complexes of Polß, showing the movements of the different domains, colored as follows: 8 kDa in green, fingers in yellow, palm in red and thumb in pink. B) Cartoon representations of the binary and ternary complex of Polλ and of the ternary complex of Pol*μ* with the same domain coloring as in A). C) Multiple sequence alignment of Polλs (L), Pol*μ*s (M) and TdTs (T) of different species. Invariant residues are shown in bold an colored red. Residues conserved among Pol*μ* ad TdTs are shown in bold. The "brooch" motif is highlighted in mauve and the residues mutated in the study are indicated with red dots. D) Cartoon representation of the Pol*μ* ternary complex with the domains colored as in A). The DNA substrate is shown in sticks, dark blue for the template strand and light blue for the primer and downstream strands. Residues subjeted to mutagenesis are shown in red sticks, residues contacting directly with them are shown in green sticks.




| | Catalytic efficiency (%primer extension) | | |
|---|---|---|---|
| | GAP filling | NHEJ comp | NHEJ incomp |
| Polμ wt | 100 | 100 | 100 |
| Y141F | 82 | 41 | 0,87 |
| Y141S | 88 | 35 | 0,31 |
| C143G | 68 | 40 | 0,41 |
| N153G | 78 | 39 | 0,91 |
| R145A | 70 | 16 | nd |
| R145K | 66 | 3 | nd |

**Figure 35. Mutagenesis of the "brooch" motif specifically affects the end-joining activity of Polμ.**
A) Gap-filling reactions (left panels) were peformed as described in Materials and Methods for the wild-type Polμ and the indicated mutants (25 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. When indicated, dNTPs were added separately at 10 nM in the presence of 2.5 mM $MgCl_2$. NHEJ reactions (right panels) were performed as described in Materials and Methods, with 200 nM Polμ and using two sets of substrates: the labeled substrates were formed by hybridization of 1TG or 1C with 1D-NHEJ, and the cold substrates, by hybridization of either 2AC or 2C with 2D-NHEJ. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, dNTPs were added separately at 10 μM in the presence of 2.5 mM $MgCl_2$. After electrophoresis, the labeled fragments were detected by autoradiography. B) EMSA was performed for the wild-type Polμ and the indicated mutants (200 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. After electrophoresis, gel was dried and the labeled fragments were detected by autoradiography. C) EMSA was performed for the wild-type Polμ and the indicated mutants (600 nM) using a 3'-protruding substrate formed by hybridization of 1TG with 1D-NHEJ. After electrophoresis, gel was dried and the labeled fragments were detected by autoradiography. E) Table and graph showing the quantification of the catalitic efficiency of the wild-type Polμ and the indicated mutants in gap-filling and NHEJ reactions.

described later.

The purified mutants were firstly analyzed on template-dependent polymerization activity assays. Gap-filling activity by these mutants was very mildly or not affected at all (both in efficiency and nucleotide insertion fidelity) when compared to wild-type Pol*μ* (Fig. 35A). As a further control, we tested by EMSA the binding of these mutants to the same gapped DNA substrate, and all displayed similar levels of DNA binding as the wild-type, and identical mobility of the enzyme/DNA binary complex formed (Fig. 35C). We next tested the activity of these mutants on complementary and non-complementary NHEJ substrates. As expected, in all of the cases NHEJ activity was reduced, but the effect reached different levels depending on the specific mutation studied. Mutants in $Tyr^{141}$, and specially the change to phenylalanine, were the ones maintaining the highest levels of activity, although they did not in any case exceed 40% of the activity of the wild-type enzyme (Fig. 35B). Mutant C143G maintained 40% of wild-type Pol*μ* NHEJ activity on compatible ends, but when tested on non-compatible ends its activity was barely detectable, even in the presence of 100-fold ddNTP concentration (Fig. 35C). A similar case was observed with mutant N153G, when tested at 100-fold ddNTP concentration compared to the wild type Pol*μ* (Fig. 35B&C). Mutations on $Arg^{145}$ showed that this residue is highly specific for NHEJ reactions: while the two mutants reached near 70% of wild-type gap-filling activity, polymerization on NHEJ substrates, either compatible or incompatible, was very low or undetectable (Fig. 35B&C). Quantification of the results obtained with the mutants in comparison with the wild type enzyme in either gapped or NHEJ substrates is shown in the graph and table in figure 18E. To test whether this inability to perform a NHEJ reaction was primarily due to defective binding of these mutants to T/D DNA substrates, we carried out EMSA experiments with the same template/downstream molecules used in the NHEJ assays. Figure 35D shows that all the mutants displayed a similar binding to a single T/D DNA substrate as the wild-type Pol*μ* (lower shifted band). That can be expected since this binding is highly dependent on the 5'-P pocket, but strikingly, none of them was able to fully reproduce the second super-shifted band that corresponds to the synapsis of two DNA ends performed by wild type Pol*μ*. Exceptionally, mutant Y141F had a moderate defect in DNA binding in agreement with its residual activity. All these results reinforce the idea that this motif is not essential to bind the T/D DNA substrate. However, it is likely that this motif is very important for the correct orientation of the 3'-protrusion. If that end is not correctly positioned, synapsis would be unstable and catalysis could not proceed. This is specially true for non-compatible ends, which depend completely on the ability of the protein to align the two 3'-overhangs, one as primer and the other as template, to achieve a correct synapsis.

**2. The BRCT domain of Pol*μ* contributes, *via* DNA binding, to its intrinsic ability of joining DNA ends**

We have shown that Pol*μ* is able to join two DNA ends in the absence of any NHEJ accessory factors such as Ku or XRCC4/LigIV (Andrade et al., 2009), to which Pol*μ* interacts *via* its BRCT domain (Bork et al., 1997), suggesting that Pol*μ* could participate in the alternative end-joining pathway that takes place without the need for the classic NHEJ factors. To further prove that this activity is inherent to the enzyme core, we tested if the Polß-like polymerization domain of Pol*μ* was able to perform DNA end-bridging and trans-directed polymerization. For this, we used a deletion mutant lacking the BRCT domain (Pol*μ*-ΔBRCT) already existing in the laboratory (Juarez, 2006; see figure 36A) and confirmed, as a control, that its 5'-P-dependent binding to gapped DNA substrates and gap-filling activity (both efficiency and fidelity) were comparable to those of the *wild-type* enzyme (Fig. 36B). DNAseI footprinting experiments on a DNA gap revealed a protection (again 5'-P-dependent) caused by Pol*μ* core (Pol*μ*-ΔBRCT) (Fig. 36E, compare lanes 4 and 5) which spans a total of 11 nucleotides of the template strand: 6 involved in base pairing the primer strand, 1 free base (the templating nucleotide in the gap; indicated with an arrow), and 4 bases paired to the downstream strand. This protection matches that obtained with Polß, which again is dependent on the 5'-P (Fig. 36E, lanes 2 *versus* 3). Moreover, those 5'-P-dependent footprints are fully compatible with the 3D structural information available for the ternary complexes of both DNA polymerases (PDB IDS: 2IHM for Pol*μ* and 1BPY for Polß). As already described in chapter 1, by using the full-length (i.e.

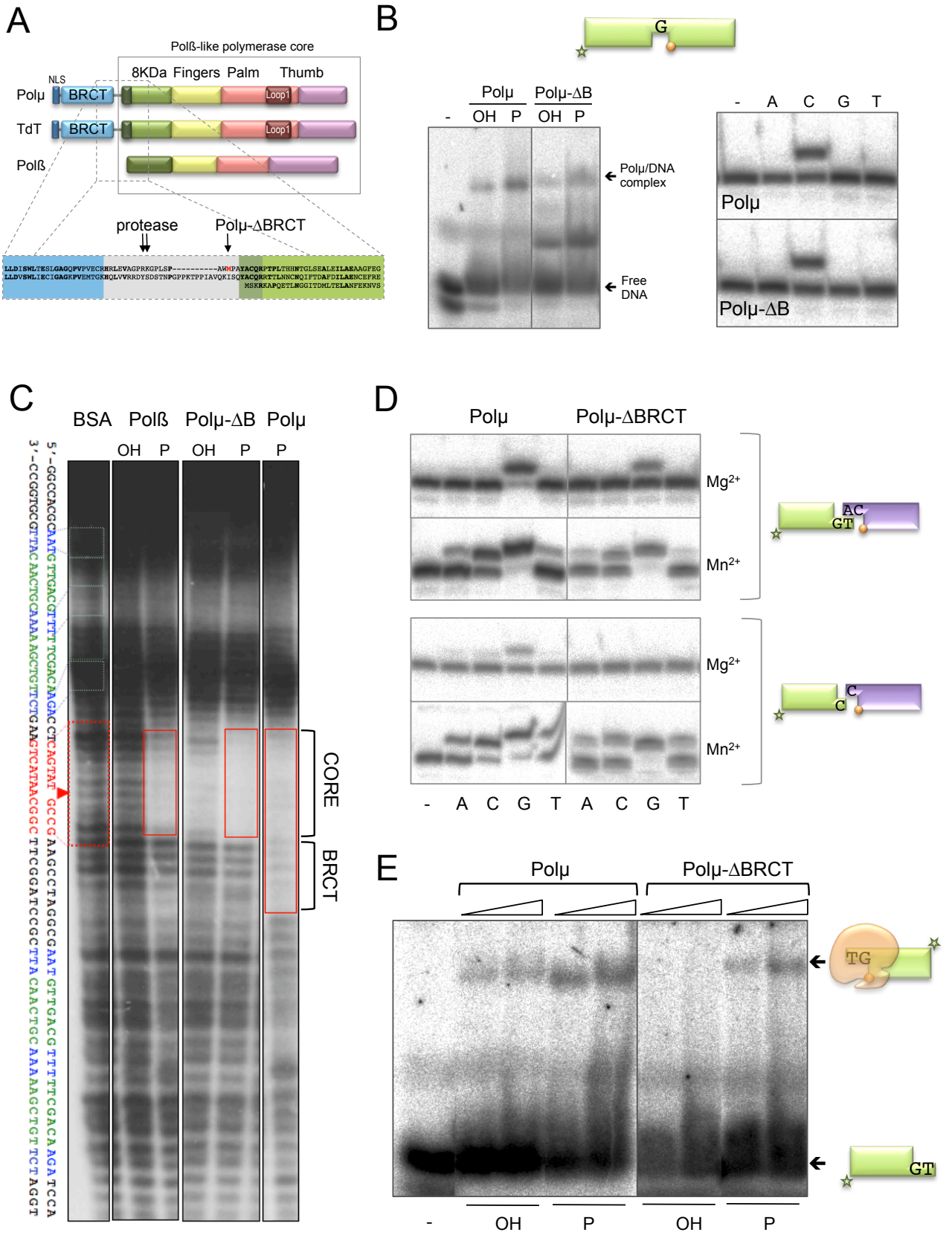



**Figure 36. The BRCT domain of Pol*μ*: implications for DNA binding.**
A) Scheme of the domain organization of Polß, Pol*μ* and TdT, and sequence alignment indicating in red the residue taken as N-terminal when constructing the deletion mutant lacking the BRCT domain. B) EMSA (left panel) was performed for the indicated proteins (200 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. After electrophoresis, gel was dried and the labeled fragments were detected by autoradiography. Gap-filling reactions (right panels) were peformed as described in Materials and Methods for the indicated proteins (25 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. When indicated, dNTPs were added separately at 10 nM in the presence of 2.5 mM $MgCl_2$. C) Footprinting assay of the wild-type Polß (5 $\mu$g) and mutant and wild-type Pol*μ* (1.5 $\mu$g). The DNA substrate used, formed by hybridizing the olignucleotides FP-T (template, labeled at its 5' end), FP-P (primer) and FP-D (downstream) may have (P) or lack (OH) a phosphate group at the 5' end of the downstream strand. D) NHEJ reactions were performed as described in Materials and Methods, with 200 nM Pol*μ* and using two sets of substrates: the labeled substrates were formed by hybridization of 1TG or 1C with 1D-NHEJ, and the cold substrates, by hybridization of either 2AC or 2C with 2D-NHEJ. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, dNTPs were added separately at 10 $\mu$M in the presence of 2.5 mM $MgCl_2$, or at 1 $\mu$M in the presence of 1 mM $MnCl_2$. E) EMSA (left panel) was performed for the indicated proteins (600 nM) using a 3'-protruding substrate containing the oligonucleotides 1TG and 1D-NHEJ.

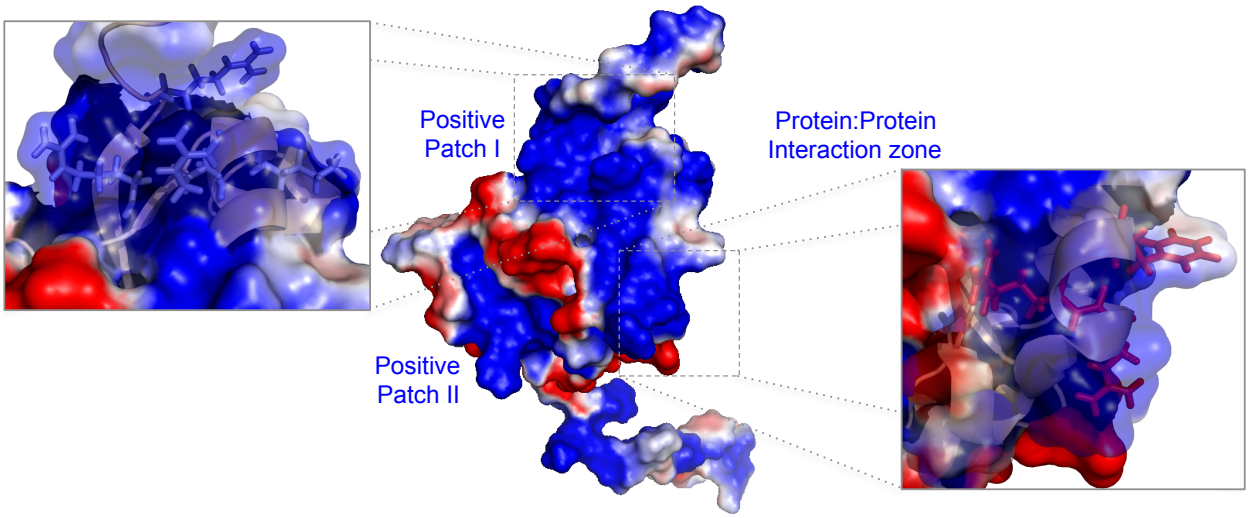
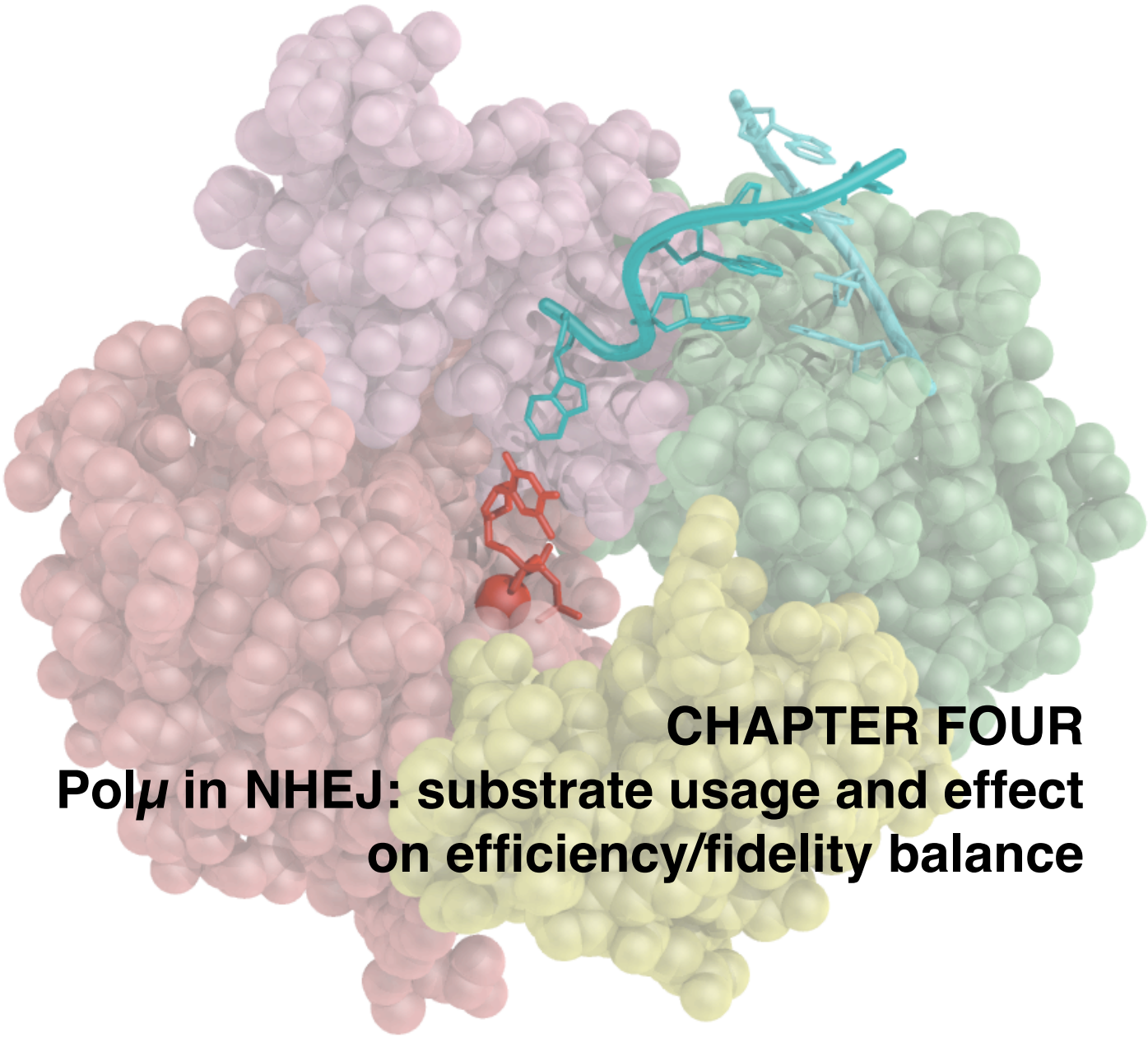



**Figure 37. BRCT domain of Pol*μ*: surfaces of interaction.**
Electrostatic surface representation of the BRCT domain of Pol*μ* (PDB ID: 2DUN). Two positive patches are indicated in the same face of the domain. The inlay to the right shows a close-up of the surface of interaction of the BRCT domain with other NHEJ factors, in the opposite face of the domain, while the inlay to the left shows a close-up of the predicted surface of interaction of the BRCT with the DNA substrate, that corresponds to positive patch I.

BRCT-containing) Pol*μ*, the footprint was further extended (6 nucleotides) towards the downstream strand, (Fig. 36E, lane 6), thus confirming the intrinsic binding of the BRCT subdomain, downstream to the core, to a gapped DNA substrate. It can be extrapolated that a similar BRCT/DNA interaction occurs during NHEJ of both complementary and non-complementary ends, being restricted to the DNA end that provides the template (the downstream side of the break), thus reinforcing the interaction already mediated by the 5'-P.

We then compared the NHEJ activity of *wild-type* and Pol*μ*-ΔBRCT mutant and the results indicated that the Pol*μ* core by itself is 3-fold less efficient than the full-length enzyme, particularly when joining incompatible ends, in which the activity is undetectable (Fig. 36C, upper panels). Manganese ions have been shown to greatly enhance the gap-filling and terminal transferase activity of Pol*μ* (Dominguez *et al.,* 2000), so we tested its capacity to reactivate the end-joining activity of the Pol*μ*-ΔBRCT mutant. In these conditions, NHEJ by the Pol*μ* core recovered activity levels similar to those of the BRCT-containing *wild-type* enzyme both at compatible and incompatible ends (Fig. 36C, lower panels). These results suggested that the BRCT domain of Pol*μ* does not play a critical role in the catalytic step of the NHEJ reaction, but could benefit NHEJ by promoting a tighter binding of Pol*μ* to the DNA substrates. Such a direct role of BRCT domains in DNA binding has been early proposed (Bork et al., 1997; Callebaut and Mornon, 1997), but direct evidence of this for BRCT-containing DNA polymerases as Polλ, yeast Pol4, Pol*μ* and TdT has not been reported yet. To corroborate this hypothesis, we tested both proteins in EMSA of short 3'-protruding DNA ends (...GT3´) having or lacking a recessed 5'-P. Whereas the full length Pol*μ* produced a stable retarded band even in the absence of a 5'-P, Pol*μ*-ΔBRCT binding to DNA was much weaker, being undetectable in the absence of a 5'-P (Fig. 36D). The BRCT domains of Pol*μ*, Polλ and TdT have been crystallized, but independently from the corresponding polymerase cores. This lack of combined structural information hamper the predictions about whether this domain could be involved in protein:DNA interactions in any of these polymerases, besides its function as a protein-protein interaction domain (Lee et al., 2004; Nick McElhinny et al., 2005). In the case of Pol*μ*, these interactions with other proteins, specially with NHEJ factors, are stablished through residues $Arg^{43}$, $Phe^{64}$ and $Leu^{50}$, as shown by DeRose et al., 2003 (see Figure 37). All these residues are located in one face of the domain, leaving two other positive patches free to interact with the DNA substrate (Fig. 37). As a result of our experimental data and helped by 3D modelling, we are able to predict the simultaneous interaction of Pol*μ* BRCT with the DNA substrate through its downstream part and with the Ku70/80 heterodimer (see Discussion).



# CHAPTER FOUR
# Pol*μ* in NHEJ: substrate usage and effect on efficiency/fidelity balance

**1. Physiological concentration of manganese ions increases Pol*μ* efficiency at no fidelity cost during NHEJ.**

As shown previously for mutant Pol*μ*-ΔBRCT, the use of manganese ions has a beneficial effect on the polymerization efficiency on NHEJ substrates, since the activity of this mutant was recovered in comparison with the very low or undetectable activity with magnesium ions. However, manganese had a negative effect on NHEJ fidelity when used at a concentration of 1 mM (Fig. 36D), a behaviour that is also true for gap-filling reactions performed by Pol*μ* and also Polλ (Angel J. Picher, 2007; Juarez, 2006). Recent work emphasized that manganese could be the natural metal activator of Polλ *in vivo*, since it is active over a wide range of $Mn^{2+}$ concentrations and is inhibited by physiological levels (10 mM) of $Mg^{2+}$ (Blanca et al., 2003). The same observations have been made for Polι which is best activated at physiological (as low as 0,1 mM) $MnCl_2$ concentrations while it is inhibited by physiological $MgCl_2$ concentrations; moreover, activation of Polι by low concentration of manganese ions improves its nucleotide insertion fidelity opposite T (Frank and Woodgate, 2007). To further address this issue, we assayed a range of $MnCl_2$ concentrations to test whether low concentration of manganese ions could modulate the fidelity and/or efficiency during gap-filling by Pol*μ*. As expected, at relatively high (100 *μ*M) $MnCl_2$ concentration, the polymerization fidelity was very poor compared to activation with 10 mM $MgCl_2$, but strikingly, fidelity progressively increased by reducing manganese concentration (40 *μ*M; 10 *μ*M) down to physiological levels (Fig. 38A): insertion of the correct nucleotide (dG) was as efficient as that obtained with 10 mM $MgCl_2$ and insertion of the other three incorrect nucleotides was barely detectable. Considering these results, we wondered if physiological concentrations of manganese ions could be also advantageous during NHEJ reactions performed by Pol*μ*. We first determined the optimal concentration of $Mg^{2+}$ or $Mn^{2+}$ required to promote peak activity of Pol*μ* on NHEJ substrates, either having compatible or incompatible ends, in the presence of metal ion concentrations ranging from 0,005 to 10 mM (Fig. 38B). In the case of $Mg^{2+}$ activation, the optimum for the two DNA substrates tested was achieved at a concentration of 2 mM, being 10 mM inhibitory (Fig. 38B, upper panels). For $Mn^{2+}$ activation, the highest level of activity was obtained at much lower concentrations, reaching the maximum between 100-200 *μ*M (Fig. 38B, lower panels). At optimal concentration of metal ions, the efficiency of the reaction was boosted by 100-200 fold when using manganese ions instead of magnesium ions, since 100 fold less dNTPs were needed to achieve even two times higher levels of incorporation when Pol*μ* was activated with $Mn^{2+}$. When we next evaluated the effect of $Mn^{2+}$ activation on the fidelity of Pol*μ*-mediated NHEJ, the results mimicked those obtained in gap-filling activity. Thus, activation by low concentrations of manganese ions allowed efficient insertion of the correct nucleotide only, while insertion of the other three was significantly reduced. This was valid for both compatible (Fig. 38C) and incompatible NHEJ substrates (Fig. 38D). Fidelity of NHEJ, measured by the ratio of insertion of the correct nucleotide relative to that of the three incorrect ones (as an average), was increased up to 3-fold (depending on the sequences of the four incompatible overhangs studied) when activated by a low (100 *μ*M) versus a high (1 mM) concentration of manganese ions (Fig. 39). These results indicate that Pol*μ* takes advantage of the use of manganese ions, at physiological concentrations, to achieve the best efficiency and fidelity when dealing with NHEJ substrates.

**2. Ribonucleotides as ultimate substrates: Insertion of ribonucleotides increases the fidelity of Pol*μ* during NHEJ.**

Ribonucleotides are highly abundant in human cells, being their levels 40- to 350-fold (depending on the dNTP/NTP pair) higher that those of deoxynucleotides in cycling cells, and 160 to 2000 fold in confluent cells (Ferraro et al., 2010; Traut, 1994). Ribonucleoside triphosphates differ from their deoxynucleotide counterparts by a single atom (oxygen) at C2' of the sugar (Fig. 40A). Thus, ribonucleotides are predicted to be inserted during DNA replication and repair at much higher frequencies than incorrect deoxynucleosides (Cavanaugh et al., 2010; Nick McElhinny et al., 2010; Nick McElhinny et al., 2010). The presence of a ribose 2'-hydroxyl group stabilizes the glycosyl bond, but makes the phosphodiester bond more

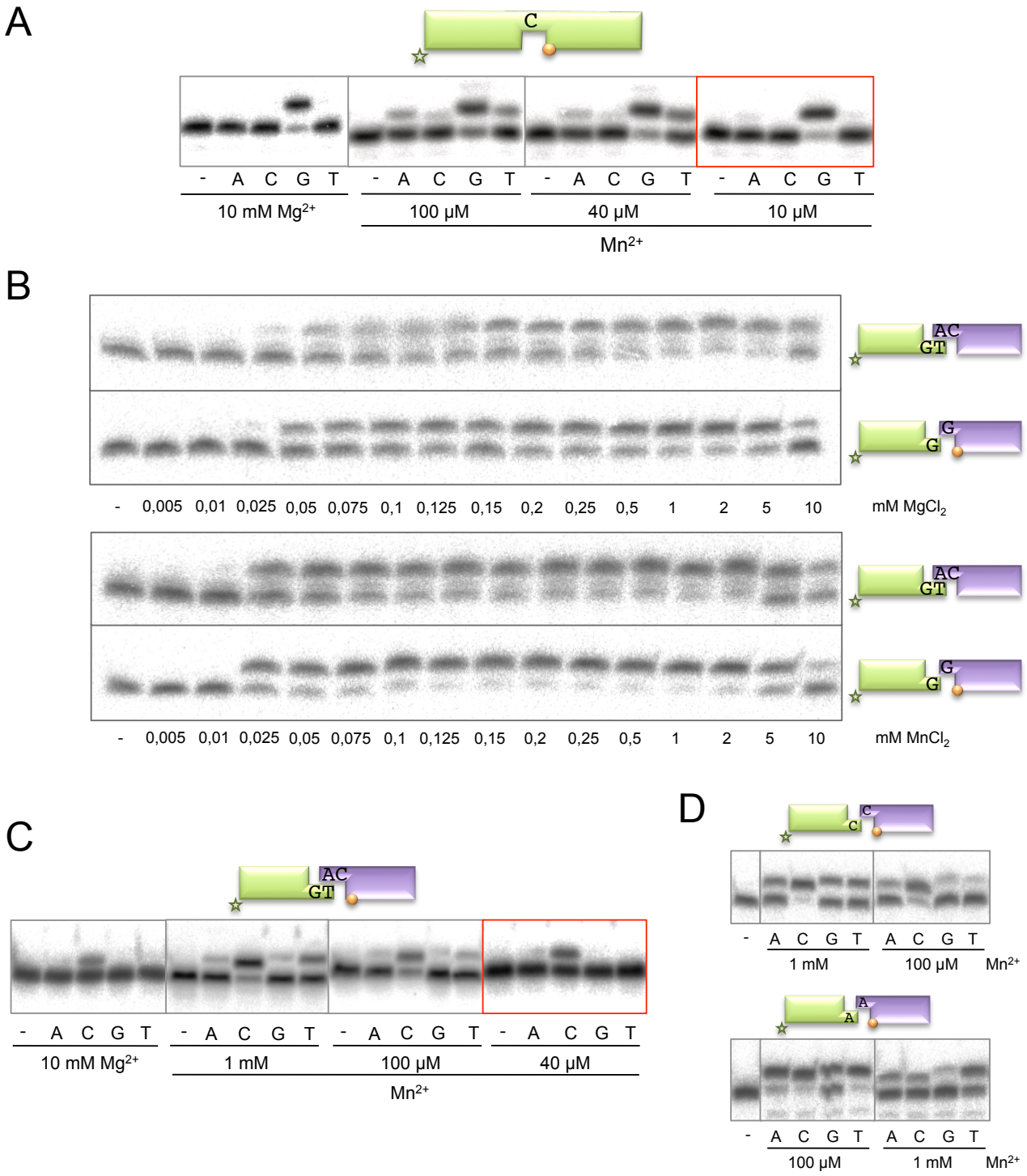



**Figure 38. $Mn^{2+}$ is the preferred activator in terms of both efficiency and fidelity during Pol$\mu$-mediated NHEJ reactions.**
A) Gap-filling reactions (right panels) were peformed as described in Materials and Methods with Pol$\mu$ (25 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. When indicated, dNTPs were added separately at 10 nM in the presence of the indicated amounts of $MgCl_2$ or $MnCl_2$. B) NHEJ reactions were performed as described in Materials and Methods, with 200 nM Pol$\mu$ and using two sets of substrates: the labeled substrates were formed by hybridization of 1TG or 1C with 1D-NHEJ, and the cold substrates, by hybridization of either 2AC or 2C with 2D-NHEJ. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, the correct ddNTP (ddG or ddC) were added at 10 $\mu$M in the presence of the indicated amounts of $MgCl_2$ or $MnCl_2$. C) NHEJ reactions were performed with 200 nM Pol$\mu$ and using one set of compatible substrates: the labeled substrate was formed by hybridization of 1TG with 1D-NHEJ, and the cold substrate, by hybridization of 2AC with 2D-NHEJ. The orange ball indicates the presence of a 5'-P group in the downstream strand of the substrate. When indicated, ddNTPs were added separately at 10 $\mu$M in the presence of the indicated amounts of $MgCl_2$ or $MnCl_2$. D) NHEJ reactions were performed with 200 nM Pol$\mu$ and using two sets of incompatible substrates: the labeled substrates were formed by hybridization of 1C or 1A with 1D-NHEJ, and the cold substrates, by hybridization of 2C or 2A with 2D-NHEJ. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, ddNTPs were added separately at 10 $\mu$M in the presence of the indicated amounts of $MnCl_2$.

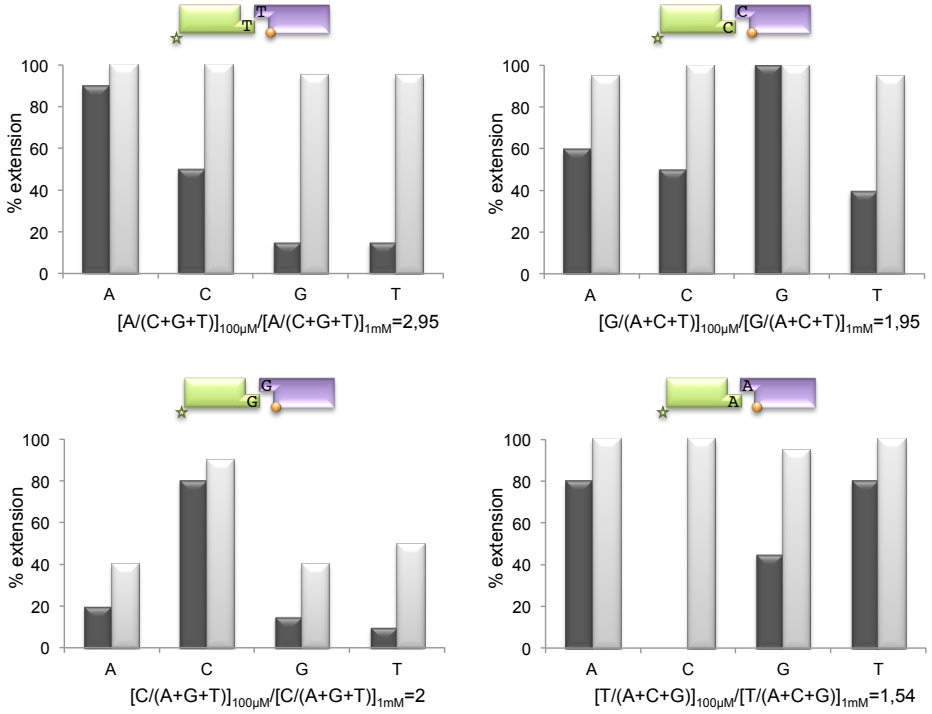



**Figure 39. Fidelity of Polμ mediated NHEJ reactions increases when using $Mn^{2+}$ at physiological concentrations.** Fidelity of NHEJ, measured by the ratio of insertion of the correct nucleotide relative to that of the three incorrect ones (as an average), was increased up to 3-fold (depending on the sequences of the four incompatible overhangs studied) when activated by a low (100 μM) versus a high (1 mM) concentration of manganese ions

susceptible to hydrolysis (Lindahl et al., 1997), thereby diminishing the chemical stability of the genome. Spontaneous or enzyme-induced strand breaks would be expected to initiate DNA repair and cellular signaling events that would also impact genome stability and cell survival.

In most instances, DNA polymerases discriminate against ribonucleotide insertion by binding them weakly and inserting them more slowly than their natural substrate. Crystallographic structures of substrate complexes of DNA polymerases from most families have suggested that a side chain could sterically interfere with binding a ribonucleoside triphosphate (rNTP) (Brown and Suo, 2011; Joyce, 1997). This side chain has been referred to as a "steric gate". In contrast, the X family polymerase Polß and Polλ have been suggested to deter insertion of a ribonucleotide using the protein backbone near the carboxyl-terminus of α-helix M, a tyrosine residue in both cases, that acts more as a "steric fence" than a "gate" (Brown et al., 2010; Pelletier et al., 1994). In the case of Polß, $Tyr^{271}$ also contacts the 3' primer terminus of the gapped substrate, and thus ribonucleotide discrimination by Polß is attributed to the loss of a contact with the primer terminus as well as a steric clash with the protein backbone (Cavanaugh et al., 2011).

As we previously reported (Ruiz et al., 2003), NTPs are valid substrates for human Polμ during gap-filling, with relatively low efficiency (about 10-fold lower) compared to dNTP insertion, but still in sharp contrast with the >1000 fold discrimination exhibited by most DNA-dependent DNA polymerases. This ability critically resides on a single residue, $Gly^{433}$, shown to play a permissive role in sugar selection of the incoming nucleotide, since its mutation to the consensus "steric fence" tyrosine present in Polß, Polλ, and even in family B DNA polymerases, as ø29 DNA polymerase (Bonnin et al., 1999), produced a strong increase in the discrimination against ribonucleotides (Ruiz et al., 2003). Another family X polymerase, Pol4 from *Schizosaccharomyces pombe*, has much smaller discrimination values, with almost equal insertion of NTPs and dNTPs in gap-filling reactions (Gonzalez-Barrera et al., 2005). There is recent evidence of the frequent usage of ribonucleotides even by replicative polymerases (Nick McElhinny et al., 2010). Strikingly, a subfamily of RNA polymerases (related to AEPs) has been involved in the NHEJ repair of DSBs in bacteria (Pitcher et al., 2007). Thus, the use of ribonucleotides as alternative substrates in such DNA repair reactions prompted us to reconsider the importance of NTP usage by Polμ, not only during gap-filling, but mainly for both the efficiency and accuracy of NHEJ reactions. Figure 40B shows that Polμ can efficiently use ribonucleotides on NHEJ substrates with a minimal microhomology of one

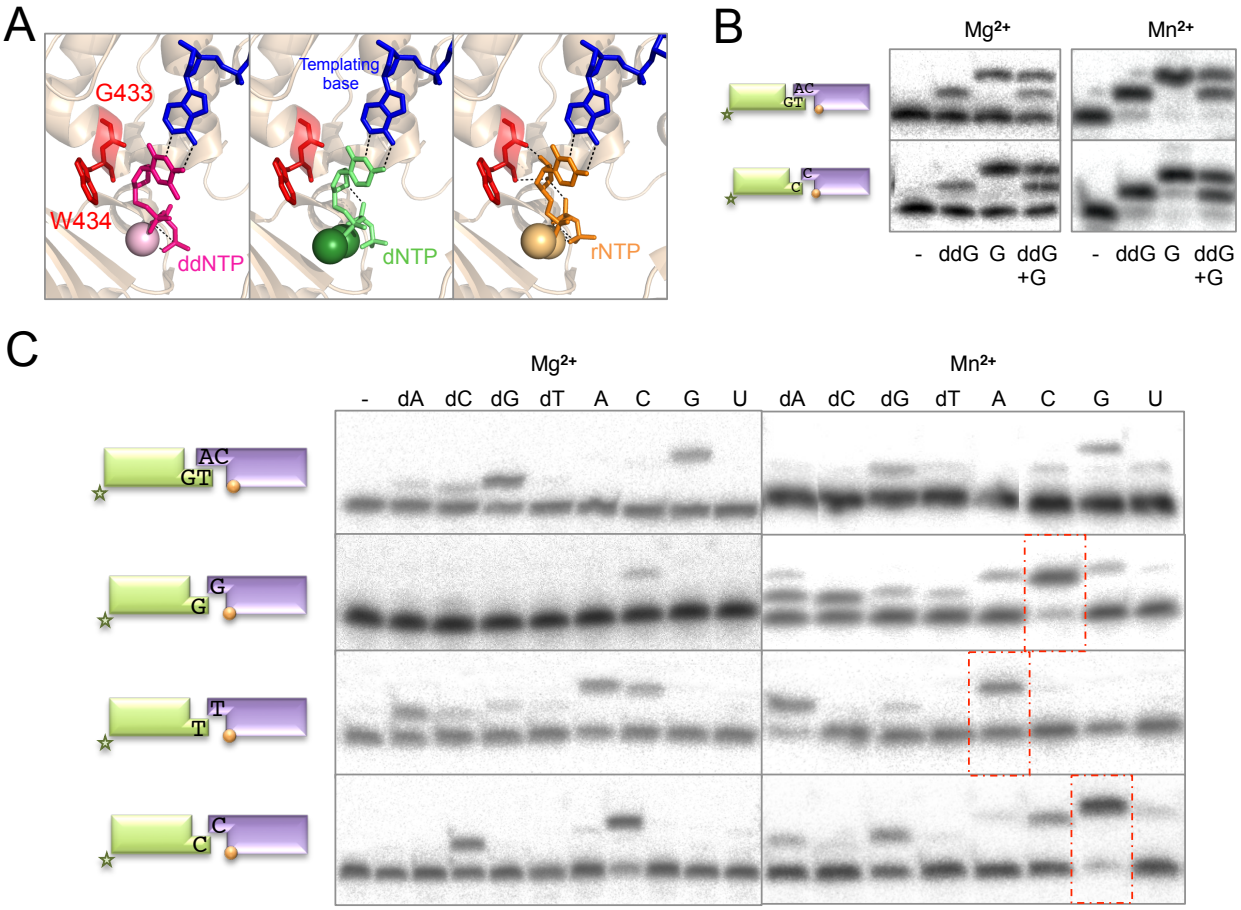



**Figure 40. Ribonucleotides as the ultimate substrates for Pol$\mu$ during NHEJ.**
A) Cartoon representation of the Pol$\mu$ ternary complex, in which the original substrates (shown in green in the middle panel: dNTP and metal ions) have been substituted with other substrates that can be used by the enzyme: ddNTP in dark pink in the left panel, NTP in orange in the rght panel. The templating base is shown in dark blue sticks and the residues forming the steric gate are shown in red sticks. B) NHEJ reactions were performed as described in Materials and Methods, with 200 nM Pol$\mu$ and using two sets of substrates: the labeled substrates were formed by hybridization of 1TG or 1C with 1D-NHEJ, and the cold substrates, by hybridization of either 2AC or 2C with 2D-NHEJ. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, the correct ddNTP (ddG, 1$\mu$M) or/and the correct NTP (rG, 10$\mu$M) were added in the presence of 2.5 mM $MgCl_2$ or 1 mM $MnCl_2$. C) NHEJ reactions were performed with 200 nM Pol$\mu$ and using four sets of substrates: the labeled substrates were formed by hybridization of 1TG, 1G, 1T or 1C with 1D-NHEJ, and the cold substrates, by hybridization of either 2AC, 2G, 2T or 2C with 2D-NHEJ. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, each of the four ddNTPs (1 $\mu$M) or NTPs (10 $\mu$M) were added in the presence of 2.5 mM $MgCl_2$ or 1 mM $MnCl_2$.

base pair (upper panels), inserting the complementary NTP (G), both when activated by either $Mg^{2+}$ (left panel) or preferentially by $Mn^{2+}$ ions (right panel). In the two cases, Pol$\mu$ showed a very low ddGTP/GTP discrimination (only 10-fold) according to the relative insertion observed, either when these two nucleotides (ddGTP at 10-fold lower concentration than GTP) were independently or simultaneously added. ddNTPs were used as substrates because X-family DNA polymerases do not discriminate between dNTPs and ddNTPs, and to limit primer extension to a single nucleotide. Very similar results were obtained when NHEJ of incompatible ends was assayed (Fig. 40B, lower panels). We then tested the fidelity of Pol$\mu$ in these contexts by comparing the insertion of the four possible bases (ribo or deoxy). In this case, we used one set of compatible and three different sets of incompatible NHEJ substrates, as shown in figure 40C. On compatible ends, preferential insertion of the correct nucleotide (dGTP or GTP) was observed when activated by either magnesium (2,5 mM) or manganese ions (100 $\mu$M). The low level of misinsertion was comparable when using ribonucleotides or deoxynucleotides, and it was maintained to a minimum in the presence of $Mn^{2+}$. Therefore, template-directed insertion of complementary ribonucleotides occurs during Pol$\mu$-driven NHEJ of compatible ends. When using non compatible ends, the outcome was slightly different as a function of the sequence of the 3'-protrusions. In all the contexts tested, efficient and predominant insertion of the correct nucleotide was observed with both metals. In all cases, however, the manganese-driven incorporation of the complementary

ribonucleotide was either the most efficient and/or accurate solution (boxed in red in Fig. 23C).
Therefore, ribonucleotides or deoxynucleotides, magnesium or manganese, can be used in alternative combinations of substrate and metal activator to achieve the optimal solution for the variety of protrusions to be handled by Pol*μ* during NHEJ.

## 3. The insertion of ribonucleotides: the consequence of a versatile active site?

As previously described, there are three Polß residues that make Van der Waals contacts with the sugar moiety (C2 and C3 carbons) of the incoming nucleotide: Tyr$^{271}$, Phe$^{272}$ and Gly$^{274}$, located at the end of α-helix M in the thumb subdomain, being Tyr$^{217}$ the "steric gate" in this polymerase (Cavanaugh et al., 2011). A counterpart of Polß Gly$^{274}$ is invariantly present in all members of the PolX family, including TdT and Pol*μ* (Fig. 41A). Interestingly, Tyr$^{271}$ and Phe$^{272}$ residues of Polß are substituted, by invariant Gly and Trp residues, respectively, in TdT and Pol*μ* from different species, and in Pol4 from *S. pombe*, all of them polymerases able to incorporate ribonucleotides. According to this, it seemed very likely that residues Gly$^{433}$ and Trp$^{434}$ of Pol*μ* could be responsible for the lack of sugar discrimination between NTPs and dNTPs. In previous studies in our laboratory (Ruiz et al., 2003), single mutations on these residues were made (Gly$^{433}$ to Tyr and Trp$^{434}$ to Phe), to restore individual consensus residues of Polß-like enzymes. Two additional mutants were obtained: W434R, to

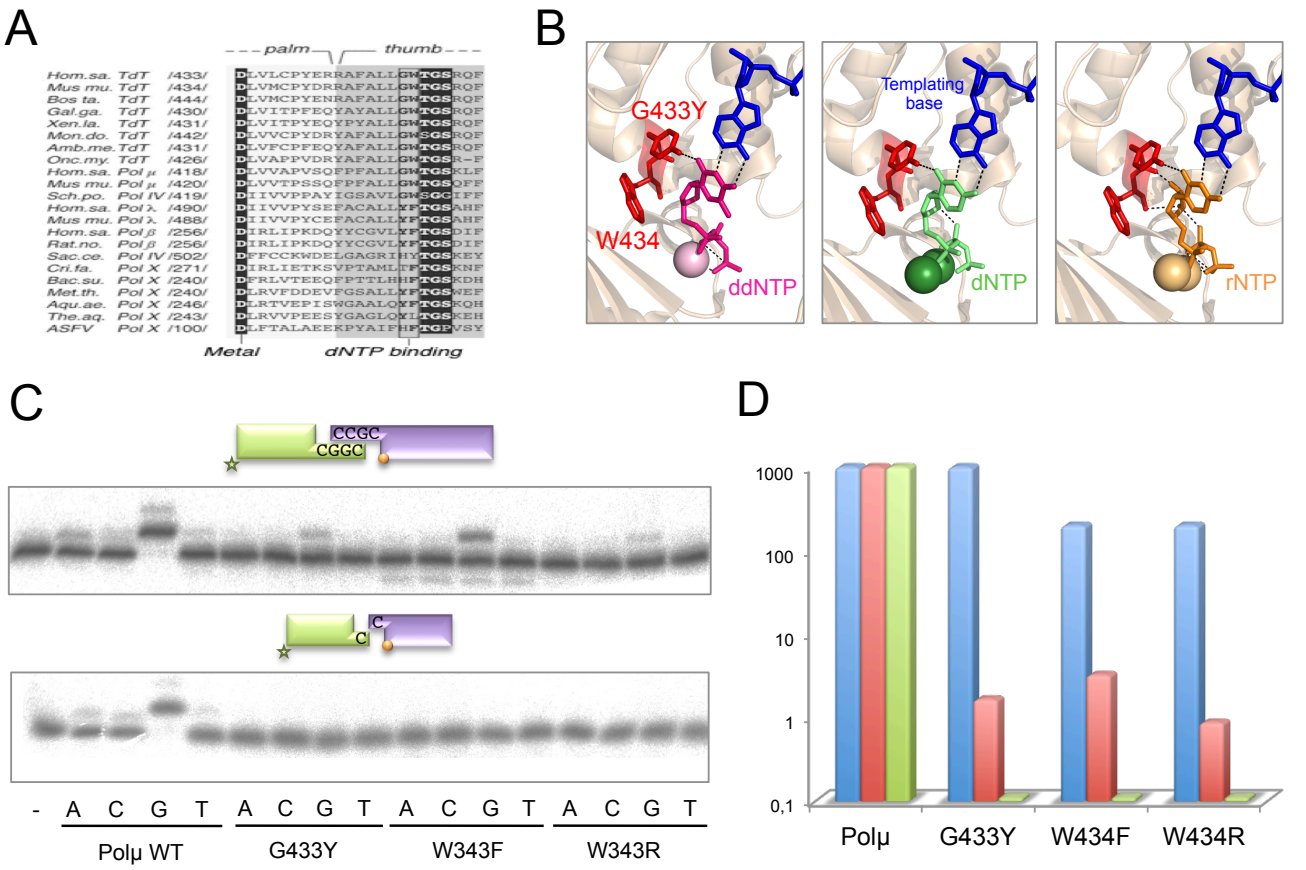



**Figure 41. Mutations in the steric gate affect insertion of ddNTPs during Pol*μ*-mediated NHEJ reactions.**
A) Multiple alignment of the amino acid region likely involved in sugar discrimination in the Pol X family. Numbers between slashes indicate the amino acid position relative to the N-terminus of each polymerase. According to Polß structural data, this region connects subdomains palm and thumb, and contains one of the three aspartates acting as metal ligands (black dot above the sequences). Invariant residues among Pol X family members are indicated in white letters over a black background. The two amino acid residues most likely implicated in sugar discrimination are indicated in bold, inside a box. Abbreviations used are: *Hom.sa., Homo sapiens, Mus mu., Mus musculus, Bos ta., Bos taurus, Gal.ga., Gallus gallus, Xen.la., Xenopus laevis, Mon.do., Monodelphis domestica, Amb.me., Ambystoma mexicanum, Onc.my., Oncorrhynchus mykiss, Sch.po., Schizosaccharomyces pombe, Rat.no., Rattus norvergicus, Sac.ce., Saccharomyces cerevisiae, Cri.fa., Crithidia fasciculata, Bac.su., Bacillus subtilis, Met.th., Methanothermobacter thermautotrophicus, Aqu.ae., Aquifex aeolicus and The.aq., Thermus aquaticus*. Taken from Ruiz *et al.*, 2003. B) Cartoon representation of the Pol*μ* ternary complex, in which the original substrates (shown in green in the middle panel: dNTP and metal ions) have been substituted with other substrates that can be used by the enzyme: ddNTP in dark pink in the left panel, NTP in orange in the rght panel. The templating base is shown in dark blue sticks and the residues forming the steric gate are shown in red sticks, one of them mutated: G433Y. C) NHEJ reactions were performed as described in Materials and Methods, with 200 nM wild-type or mutant versions of Pol*μ* and two sets of substrates: the labeled substrates were formed by hybridization of D3-C with D1 or 1C with 1D-NHEJ, and the cold substrates, by hybridization of either D4-C with D2 or 2C with 2D-NHEJ. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, each of the four ddNTPs (10 *μ*M) were added in the presence of 2.5 mM $MgCl_2$. D) Quantification of the catalytic efficiency of the indicated enzymes on gap-filling (blue bars), NHEJ of compatible ends (red bars) and of incompatible ends (green bars), in a logarithmic scale.

evaluate the effect of a non-conservative change at this position, and the double mutant G433Y/W434F, to fully restore the Polß consensus. Mutant W434F had a polymerization activity similar to that of the wild-type protein, as assessed by dNTP incorporation, but the polymerization of ribonucleotides occurred at lower rate compared to the wild-type. In agreement with the conservation of an aromatic residue in most family X polymerases, the catalytic activity of mutant W434R decreased about 5-fold compared to the wild-type protein, with an additional decrease in the insertion of ribonucleotides. Strikingly, the ability to incorporate NTPs drastically dropped in the Pol*μ* mutant G433Y, but this mutation also resulted in a decrease in the incorporation of dNTPs, which suggests that this amino acid residue is also important for the optimal catalytic activity of the enzyme, perhaps participating in the formation of a catalytically competent ternary complex (Fig. 41B). The double mutant protein G433Y/W434F had an almost null polymerization activity with both dNTPs and NTPs, indicating the essential role of these amino acid residues for an optimal polymerization activity of Pol*μ*. The negative effect of this mutations, specially the change of the glycine for a larger residue, in the general polymerization performed by Pol*μ* may point to a unique requirement of an open and spacious active site, that leads not only to the insertion of ribonucleotides and other nucleotide analogues, but also to higher misincorporation rates and, most specially, to the ability to bind and connect two distinct DNA ends that may contain flaps, mismatches, damaged bases, etc.

To test this hypothesis, we studied if the different competence of these mutants in inserting NTPs would specifically affect the efficiency and fidelity of Pol*μ*-driven NHEJ reactions. As shown in figure 41C, the polymerization levels of the mutants were greatly affected (compatible ends) or undetectable (non-compatible ends). The graph in figure 41E shows the quantification of the +1 incorporation of the correct deoxynucleotide on a DNA gap (blue)compared with compatible (red) or incompatible (green) NHEJ substrates, on a logarithmic scale. The effect of the mutations in this zone of the protein is sensibly larger when confronting the most difficult reaction of bridging two DNA ends and perform trans-directed polymerization.

## 4. Nucleotides and cofactors used *in vivo* for NHEJ.

### 4.1 DNA polymerases present at HeLa extracts.

The NHEJ pathway of DNA DSBs repair is considered to be error-prone (Lukacsovich et al., 1994; Rouet et al., 1994) but, nevertheless, it can often restore the original sequence at DSB sites, even for DSBs with damaged ends and fragmented nucleotides (Chen et al., 2001), as we have already shown (chapter 3). This repair is accomplished by a mechanism involving alignment of residual complementarities in 3' or 5' overhangs, fill in of any remaining gaps, and ligation (Pfeiffer et al., 1994; Roth and Wilson, 1986). This process can be mimicked *in vitro* using whole-cell extracts of human cells that will provide, in addition to the NHEJ DNA polymerases, the required Ku, XRCC4, and DNA ligase IV (Chen et al., 2001; Feldmann et al., 2000). The XRCC4-DNA ligase IV complex (XRCC4/LigIV) is required not only for the final ligation, but for the gap-filling step as well (Lee et al., 2003). The gap-filling polymerase activity has not been definitively identified, but there is circumstantial evidence that it may be either Pol*μ* (Mahajan et al., 2002; Zhang et al., 2001) or Polλ (Bebenek et al., 2003).

To further understand the role of Pol*μ* and its use of ribonucleotides/manganese in the NHEJ repair pathway *in vivo,* we carried out different experiments to test various DNA repair activities using nuclear cell extracts derived from the human HeLa tumoral cell line. This type of test is based on previous studies performed in collaboration with the group of Dr. Povirk (Virginia Univ., USA), which were focused on studying the role of Polλ during NHEJ by using the same type of Hela extracts, and substrates which, once joined, would display a complementarity of several base pairs. Those results suggested that nuclear extracts of human HeLa cells, supplemented with recombinant XRCC4/LigIV, were capable of accurately rejoining double-strand break substrates with a 1- or 2-base gap, and the gap-filling step was dependent on XRCC4/ligase IV (Lee et al., 2004). To determine what polymerase was responsible for gap filling, end joining was examined in the presence of polyclonal antibodies against each of two prime candidate enzymes, Polλ and Pol*μ*, both of which were present in the extracts. The results suggested that Polλ is the primary gap-filling polymerase for accurate NHEJ of complementary ends, and that the

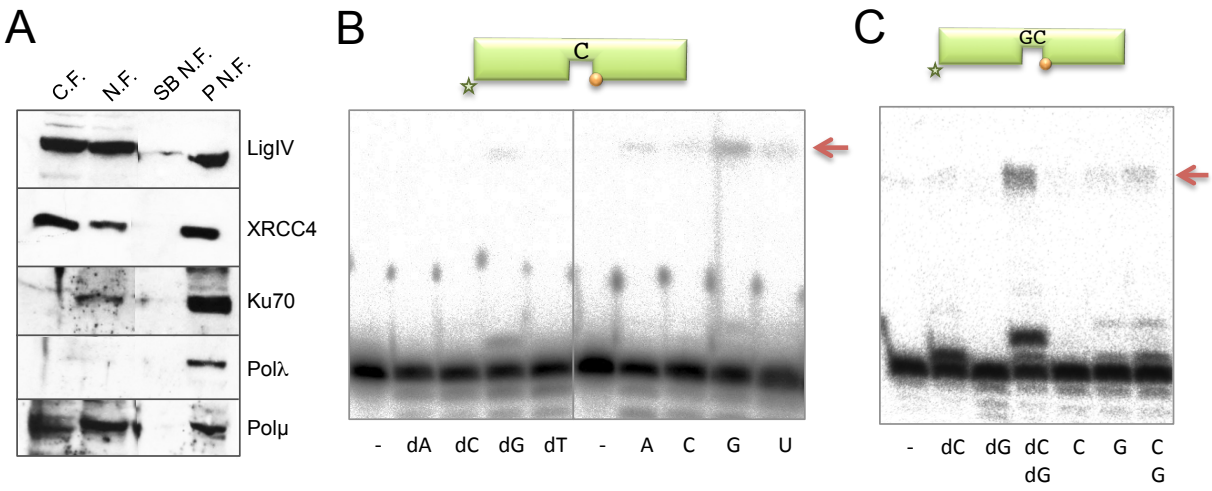



**Figure 42. Use of ribonucleotides during repair reactions by HeLa nuclear extracts.**
A) Western blot analysis showing the presence of the different NHEJ factors (Ligase IV, XRCC4, Ku70, Polλ or Pol*μ*) in the different steps of the purification of the HeLa nuclear extracts: C.F. (cytosolic fraction), N.F (nuclear fraction), SB N.F. (supernatant of the nuclear fraction), P N.F. (pellet of the nuclear fraction - this is the fraction used in the assays).The antibodies used were polyclonal α-Pol*μ* and α-Polλ (own source) and mouse monoclonal antibodies (sc9033, sc5606 and sc11750, Santa Cruz Biotechnologies). B) Gap-filling reactions were peformed as described in Materials and Methods with HeLa nuclear extracts (30 *μ*g) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. When indicated, dNTPs or NTPs were added separately at 100 *μ*M in the presence of the indicated amounts of 1 mM $MnCl_2$. After 2h of incubation, reactions were stopped and loaded on 20% PAGE - 8 M urea gels. After electrophoresis, labeled fragments were detected by autoradiography. +1 incorporation and ligation (red arrow) products could be detected. C) Gap-filling reactions were carried out as in B), but using a 2 nt gapped substrate containing the oligonucleotides P15, T17 and DG2-P.

BRCT domain is required for this activity (Lee et al., 2004). Instead of using the plasmid-based substrates utilized in Dr. Povirk's study, we decided to use the oligo-based substrates that we had already tested with the purified protein to maintain coherence with our previous results.

We tested the HeLa nuclear extracts and the discarded cytoplasmic fractions for the presence of NHEJ accessory factors needed for the reaction. Western-blotting showed that the NHEJ factors XRCC4, Ligase IV and Ku70 were enriched in the nuclear fractions obtained (Fig. 42A). To determine whether Pol*μ* and Polλ were also present in our HeLa nuclear extracts, the same blots were probed with antibodies to each polymerase. Although both antibodies detected multiple species in the extract, in each case there was a prominent band that comigrated with the corresponding purified recombinant protein at approximately the expected molecular mass (Fig. 42A). Densitometric comparison of chemiluminescence signals produced by the extracts and various quantities of each recombinant polymerase suggested that 30 *μ*g of nuclear extract (the amount used in each assay) contained 10 ng of Pol*μ* and 20 ng of Polλ, and that a single HeLa nucleus contains about 60.000 molecules of Pol*μ* and 100.000 of Polλ.

To study the capacity of the extracts of completing one of the steps of a NHEJ reaction we used a 1-nt gapped substrate, and after incubation for 6 hours at 37ºC, we detected the polymerization reaction only with the correct deoxynucleotide (dG) dictated by the templating base (Fig. 42B). We were also able to detect ligation (after gap-filling/dG incorporation) by the formation of a product of 34 nts, (red arrow). Both steps of the repair reaction occurred when activated either by magnesium (first panel), or by manganese ions (second panel). Strikingly, we were also able to detect efficient and faithful polymerization with ribonucleotides (Fig. 25B, third and fourth panels), an activity specific to Pol*μ* on this kind of substrates, and the subsequent ligation. This ligation reaction was even more efficiently performed by the extract in the presence of NTPs than dNTPs, which could account for the known preference of LigIV of ligating substrates containing a 3'-terminal NTP.

Looking for some activity that could be more specifically attributed to Pol*μ*, we next used a 2-nt gap in which Pol*μ*, but not Polß or Polλ, preferentially dislocates the template to incorporate the nucleotide complementary to the second templating base (Juarez, 2006). We performed this experiment in the presence of $Mg^{2+}$ as a cofactor and with both deoxy and ribonucleotides as

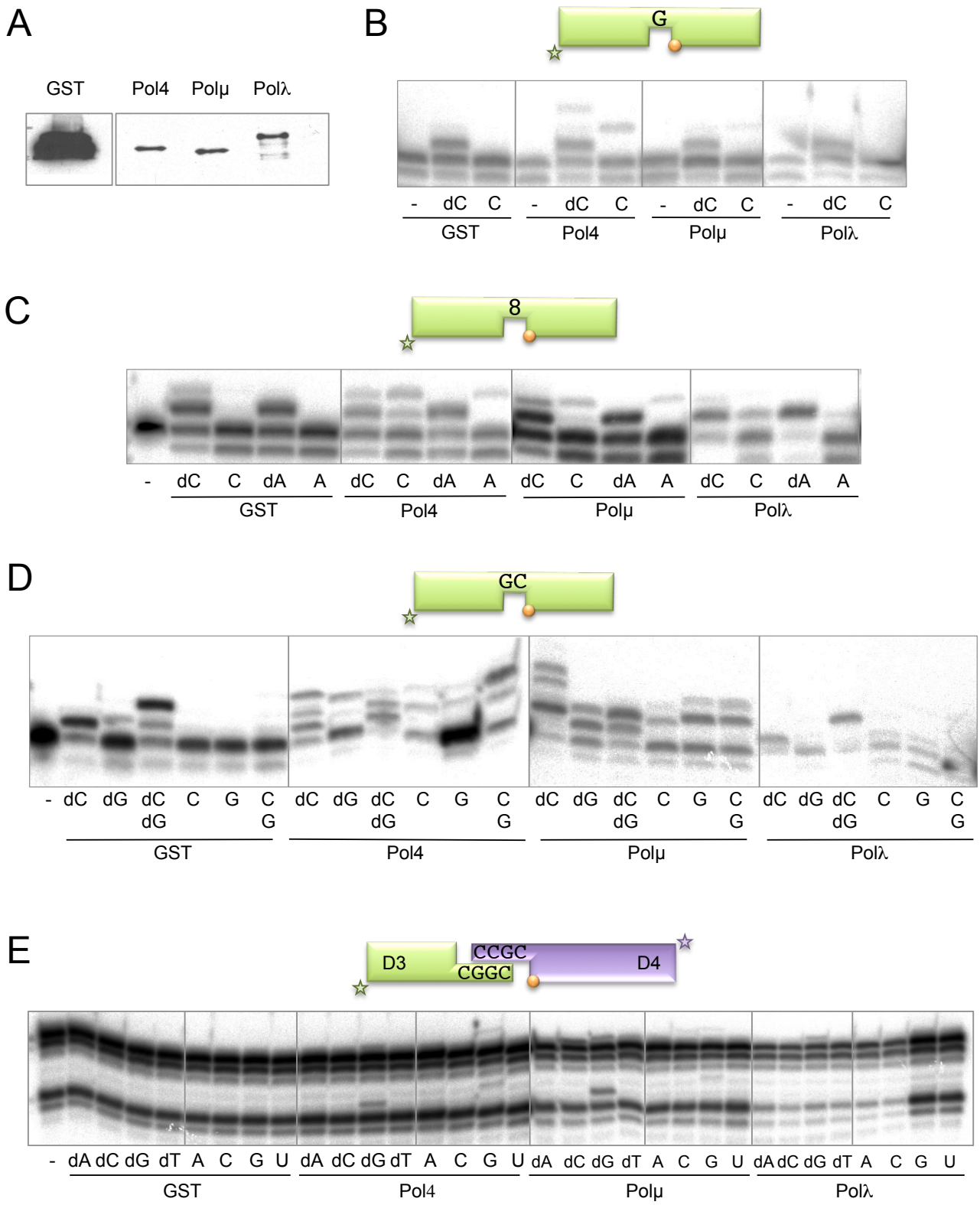



**Figure 43. Use of ribonucleotides during repair reactions by *S. pombe* extracts over-expressing Pol4, Polλ or Polμ.**
A) Western blot analysis showing the presence of the over-expressed GST-tagged proteins in the yeast extracts. B) Gap-filling reactions were peformed as described in Materials and Methods with S. pombe extracts (30 μg) using a gapped substrate containing the oligonucleotides SP1C, T13G and DG1-P. When indicated, dNTPs or NTPs were added separately at 100 μM in the presence of the indicated amounts of 1 mM $MnCl_2$. After 2h of incubation, reactions were stopped and loaded on 20% PAGE - 8 M urea gels. After electrophoresis, labeled fragments were detected by autoradiography. C) Gap-filling reactions were carried out as in B), but using a gapped substrate with a 8oxoG as the templating base, formed with the oligonucleotides pBER, dBER and T8. D) Gap-filling reactions were carried out as in B), but using a 2 nt gapped substrate containing the oligonucleotides P15, T17 and DG2-P. E) NHEJ reactions were performed as described in Materials and Methods, with 30 μg of *S. pombe* extracts and one set of substrates: the short substrate (green) was formed by hybridization of D3-C with D1 and the long substrate (mauve) by hybridization of D4-C with D2. The orange ball indicates the presence of a 5'-P group in the downstream strand of the substrate. When indicated, each of the four dNTPs or NTPs (100 μM) were added in the presence of 1 mM $MnCl_2$.

substrates. In the presence of dNTPs, the polymerase involved has a preference to insert the first templated nucleotide (dCTP), it does not anticipate the insertion of dGTP, the nucleotide complementary to the second templating base (Fig. 42C, compare lanes 2 and 3), and efficiently fills the 2-nt gap when given both dCTP and dGTP at the same time. The formation of a 34-mer, product of a very efficient ligation reaction, could be observed when the gap was correctly filled, but some level of ligation was also observed when the incorporation of only the first nucleotide was achieved. In the case of the ribonucleotides, no reaction was observed with CTP, complementary to the first templating base, but on the other hand, and even though the reaction levels were lower than with deoxynucleotides, the incorporation of GTP, complementary to the second templating base of the gap, could be clearly observed; that indicates that the first templating base is being dislocated, thus pointing at Pol*μ* as the main author of this incorporation. When the two ribonucleotides CTP and GTP were present, the reaction pattern was the same as when only GTP was supplied, again a typical Pol*μ* behaviour. In the cases in which incorporation was observed, ligation products were also formed (red arrow), implying the formation of a ligated product lacking 1 nucleotide.

### 4.2 *S. pombe* extracts over-expressing SpPol4, Polλ or Pol*μ*.

In contrast to mammals, budding and fission yeasts have only one DNA PolX enzyme (Burgers et al., 2001). Whereas *Saccharomyces cerevisiae Sc*Pol4 is closely related to Polλ (25% amino acid identity to the Polλ core), the DNA PolX from the fission yeast *S. pombe*, *Sp*Pol4, is more closely related to Pol*μ* than to Polλ (27% versus 24% identical core residues, respectively). *Sc*Pol4 was the first DNA PolX shown to play a role in NHEJ (Wilson and Lieber, 1999), and its capacities during yeast NHEJ have been studied in great detail by Daley et al., 2005. In this work the authors determined that *Sc*Pol4 was not needed during the repair of DSBs with 5'-overhangs, while repair of DSBs with 3'-overhangs and a gap on each strand strongly depend on the X family polymerase. When the synapsis of two 3'-overhangs contained a gap on only one strand, and ligation could proceed immediately in the other strand, *Sc*Pol4 was no longer needed for repair. *Sc*Pol4 was required at 3'-overhangs of all lengths within the NHEJ-dependent range but was dispensable outside of this range, indicating that this polymerase is specific to NHEJ. In agreement with all that data that implicate *Sc*Pol4 in the NHEJ pathway, a direct interaction of the BRCT domain of ScPol4 with the Lig4/Lif1 complex (homologue to human XRCC4/LigIV) has been demonstrated (Tseng and Tomkinson, 2002), as well as a physical and functional interaction of Rad27 (homologue to human FEN-1) with both *Sc*Pol4 and Dnl4/Lif1 (Tseng and Tomkinson, 2004). In the case of *Sp*Pol4, its DNA polymerization properties and the presence of a dRP lyase activity support a role of this DNA polymerase in both NHEJ and BER reactions. In spite of the closer similarity to Pol*μ*, this enzyme combines Polß, Pol*μ* and Polλ properties (Gonzalez-Barrera et al., 2005). Similarly to Pol*μ*, *Sp*Pol4 can incorporate a ribonucleotide (NTP) into a primer DNA. However, it is not responsible for the 1–2 ribonucleotides proposed to be present at the mating-type locus and those necessary for mating-type switching (Gonzalez-Barrera et al., 2005). Unlike Pol*μ*, *Sp*Pol4 lacks terminal deoxynucleotidyl transferase activity and realigns the primer *terminus* to alternative template bases only under certain sequence contexts and, therefore, it is less error prone than Pol*μ* (Gonzalez-Barrera et al., 2005). Finally, mammalian family X polymerases have been shown to differentially complement a *pol4* mutation in *S. cerevisiae*, depending on the joint structure, demonstrating that these polymerases can participate in yeast NHEJ but with distinct properties. In the light of all this evidence, we decided to use *S. pombe* as a model organism since it provides us with a simplified system to test the usage of substrates by *Sp*Pol4 with a more physiological approach, using whole cell extracts instead of purified recombinant proteins, and, more interestingly, of Pol*μ* and Polλ over-expressed in this system. In order to eliminate the possible background of an endogenous *Sp*Pol4 activity, we used a *S. pombe* mutant strain lacking the *pol4* gene, *ΔPol4*. To achieve overexpression of the polymerases of interest we used an expression vector under the control of the *nmt1* promoter, that upon elimination of thiamine from the culture medium would initiate expression of the GTS tagged versions of either SpPol4, human Pol*μ* or human Polλ. As a negative control, extracts from cells over-

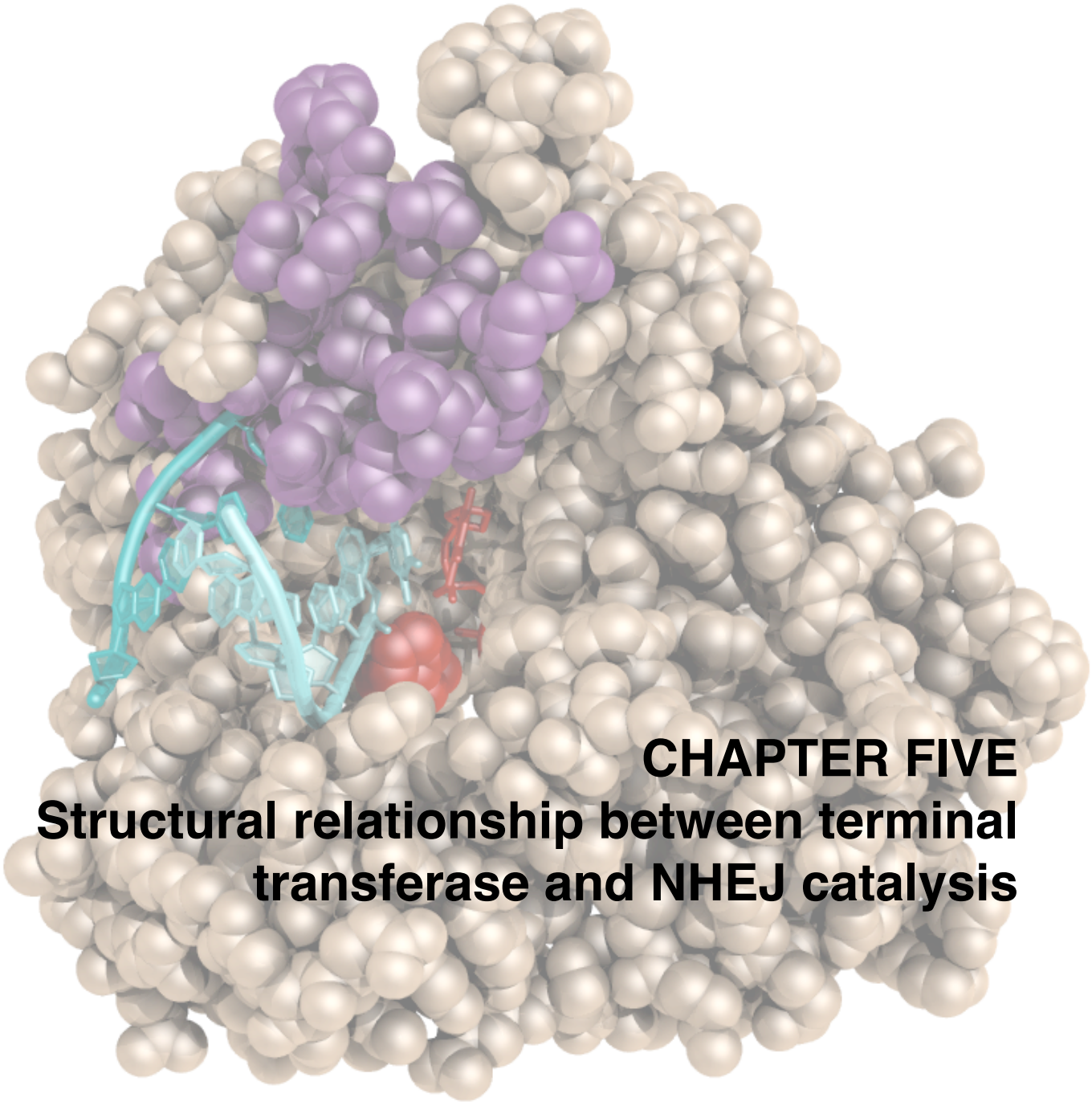
expressing GST from the empty vector were also assayed. Once the strains were obtained, we determined that the levels of over-expression of the three polymerases in the whole cell extracts were very similar, as assessed by western-blotting experiments with α-GST monoclonal antibody (Fig. 43A).

Polymerase activity of the extracts was first assayed in the presence of a 1-nt gapped DNA substrate (Fig. 43B), and our results show that the GST extract displays a very efficient gap-filling activity when manganese ions and the correct deoxynucleotide are supplied, while the endogenous polymerase activity responsible for this reaction was not able to use the correct ribonucleotide as a substrate (Fig. 43B, lane 3). This feature is important since, as it has been already shown (Gonzalez-Barrera et al., 2005; Ruiz et al., 2003), Pol*μ* and SpPol4 are able to make use of NTPs, thus making it a specific activity, ascribable only to these polymerases in this context. For instance, when either of these two polymerases was over-expressed, incorporation of the correct rNTP was observed (Fig. 43B lanes 6 and 9), but not in the case of Polλ, as expected (lane 12).

To test a substrate more related to the DNA repair pathways in which these polymerases are involved, we next used a 1-nt gapped DNA substrate in which the templating base is 8oxoG, an oxidized damaged base (Fig. 43C). We supplied the extracts with the deoxy and ribo versions of adenine and cytosine, and we can observe that similar levels of the non-mutagenic and mutagenic incorporation of dC and dA are obtained in the case of the four extracts, and that the use of ribonucleotides, both rA and rC, is again restricted to the *Sp*Pol4 and Pol*μ* extracts. In the case of Polλ, extension of the primer with deoxynucleotides is much more efficient than that observed in the GST extract, and this could account for a specific role of Polλ in this kind of DNA intermediates, that mimick those of the repair pathway of the 8oxoG:dAMP mismatch, in which Polλ has been implicated (van Loon and Hubscher, 2009).

The substrate that would help us visualize the activity most specifically related with Pol*μ*, the dislocation of the template, is the 2-nt gapped substrate. The GST and Polλ extracts, as expected, were not able to efficiently dislocate the template and thus incorporated preferentially the deoxynucleotide complementary to the first templating base (Fig. 43D, lanes 2 and 20). When the two nucleotides are supplied, complete filling of the gap is achieved (lanes 4 and 22). Again, ribonucleotides are not valid substrates for the polymerases implicated in this gap-filling activity. In the case of the SpPol4 and Pol*μ* extracts, on the other hand, dislocation was very efficiently achieved, both with deoxy and ribonucleotides, indicating a replacement of the endogenous polymerase activity on this substrate (Fig. 43D, second and third panels).

Finally, to check the NHEJ activity displayed by the extracts we used a set of NHEJ substrates, similar to those used in section 1, formed by two double stranded and 3-protruding molecules, that in this case display a complementarity of 3 C:G base pairs when joined. The length of the two molecules is different enough so that both can be labelled at the same time, to allow detection of the nucleotide incorporation achieved in the two gaps at the same time. In this case, the GST extract displayed no activity at all (Fig. 43E, first panel), showing that the endogenous polymerase activity, although capable of filling a gap, is not able to join to DNA molecules. The Polλ extract showed a limited ability to join these DNA ends and inserted the correctly template deoxynucleotide only on one side of the synapsis (Fig. 43E, fourth panel). The *Sp*Pol4 and Pol*μ* extracts, on the other hand, were able to efficiently and faithfully polymerize on both sides of the synapsis, and with both deoxyribonucleotides and ribonucleotides (Fig. 43E, second and third panels).



# CHAPTER FIVE
# Structural relationship between terminal transferase and NHEJ catalysis

**1. Is terminal transferase operating during NHEJ of incompatible ends? Importance of a recessive 5'-P.**

It has been recently reported that Pol*μ*´s optimal end-joining of incompatible ends occurs at minimal (1-2 nt) overhangs; on longer protrusions, a single polymerase unit cannot maintain bridging of the two ends, and the yield of the reaction is lower, exclusively reflecting terminal transferase activity (Davis et al., 2008). Since recessive 5' DNA ends must contain a 5'-P group to allow ligation and repair of the DNA break, and family X polymerases contain an 8 kDa domain that interacts directly with the 5'-P group of the downstream strand, we determined if the presence of a 5'-P group modulates Pol*μ*´s terminal transferase activity, perhaps favoring accurate NHEJ of short incompatible ends. As shown in Figure 44A (left), in the presence of a 5'-OH group the terminal transferase activity of human Pol*μ* extends the 1nt-protruding end with any of the 4 ddNTPs, ddCTP being the most efficient. However, the presence of a 5'-P group (right) alters this specificity: incorporation of ddTTP, ddCTP and ddGTP is significantly reduced, but incorporation of ddATP is largely enhanced. These results can be explained as indicated in Figure 44B: if a 5'-P is not present (scheme 1), Pol*μ* inserts untemplated nucleotides (N) by terminal transferase; if a 5'-P group is present (scheme 2), the preferred insertion of ddATP indicates that a single Pol*μ* molecule was able to bridge two incompatible DNA ends, and copy the templating base (T) provided by a second end, thus achieving a precise NHEJ reaction, as it was shown to occur at different incompatible ends also when Pol*μ* was in the presence of NHEJ core factors

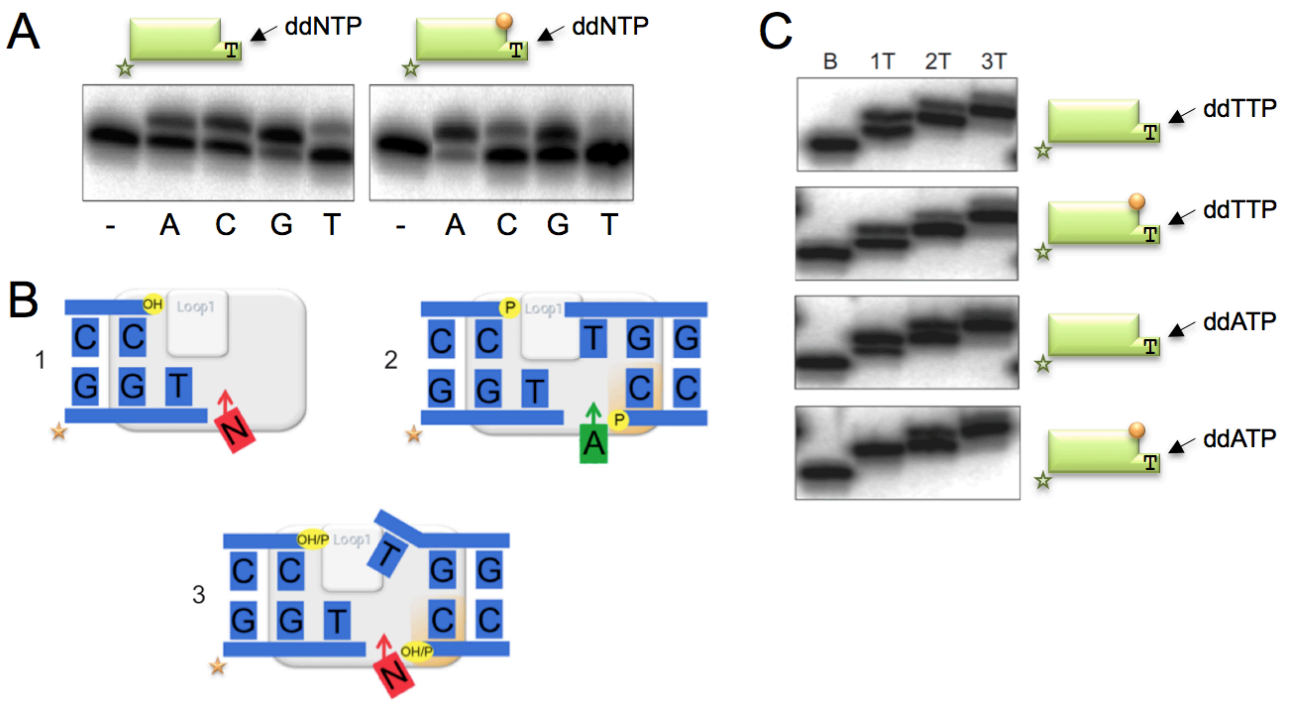



**Figure 44. Terminal transferase vs. NHEJ of incompatible ends.**
A) Single nucleotide extension of a 3'-protruding (1T) dsDNA substrate (200 nM), either lacking (left panel) or having (right panel) a 5'-recessive P. The assay was performed as described in Materials and Methods, in the presence of 1 mM MnCl2, 100 □M of the indicated ddNTP, and 200 nM of Pol*μ* wt. After incubation for 30 min at 30 °C, +1 extension of the 5'-P-labeled oligonucleotide was analyzed by 8 M urea-20% PAGE and autoradiography. B) The observed nucleotide specificity can be the consequence of either untemplated polymerization (terminal transferase; scheme 1) or DNA-directed additions (dA insertion templated in trans), thus achieving precise joining of two DNA ends (scheme 2). Scheme 3 shows a hybrid situation in which the joining reaction is untemplated, as the 3'-protruding template base provided in trans is not in a proper register. The gray rectangle represents a cartoon of Pol*μ*, in which the specific loop 1, involved in both terminal transferase and connec- tivity during NHEJ of incompatible ends, is highlighted. The orange area in Pol*μ* represents the 8-kDa domain, specifically involved in 5'-P recognition. C) Switching from terminal transferase to NHEJ mode, dependent on the presence of a 5P group only occurs with 1nt protruding substrates. The assay was performed as described in Materials and Methods, using either blunt or 3'-protruding (1T, 2T, and 3T) dsDNA substrates (200 nM), with or without a recessive 5'-P, as indicated, in the presence of 1 mM MnCl2, 100 M of the indicated ddNTP, and 200 nM of Pol wt. After incubation for 30 min at 30ºC, the +1 extension of the 5'-P-labeled oligonucleotide was analyzed by 8 M urea-20% PAGE and autoradiography

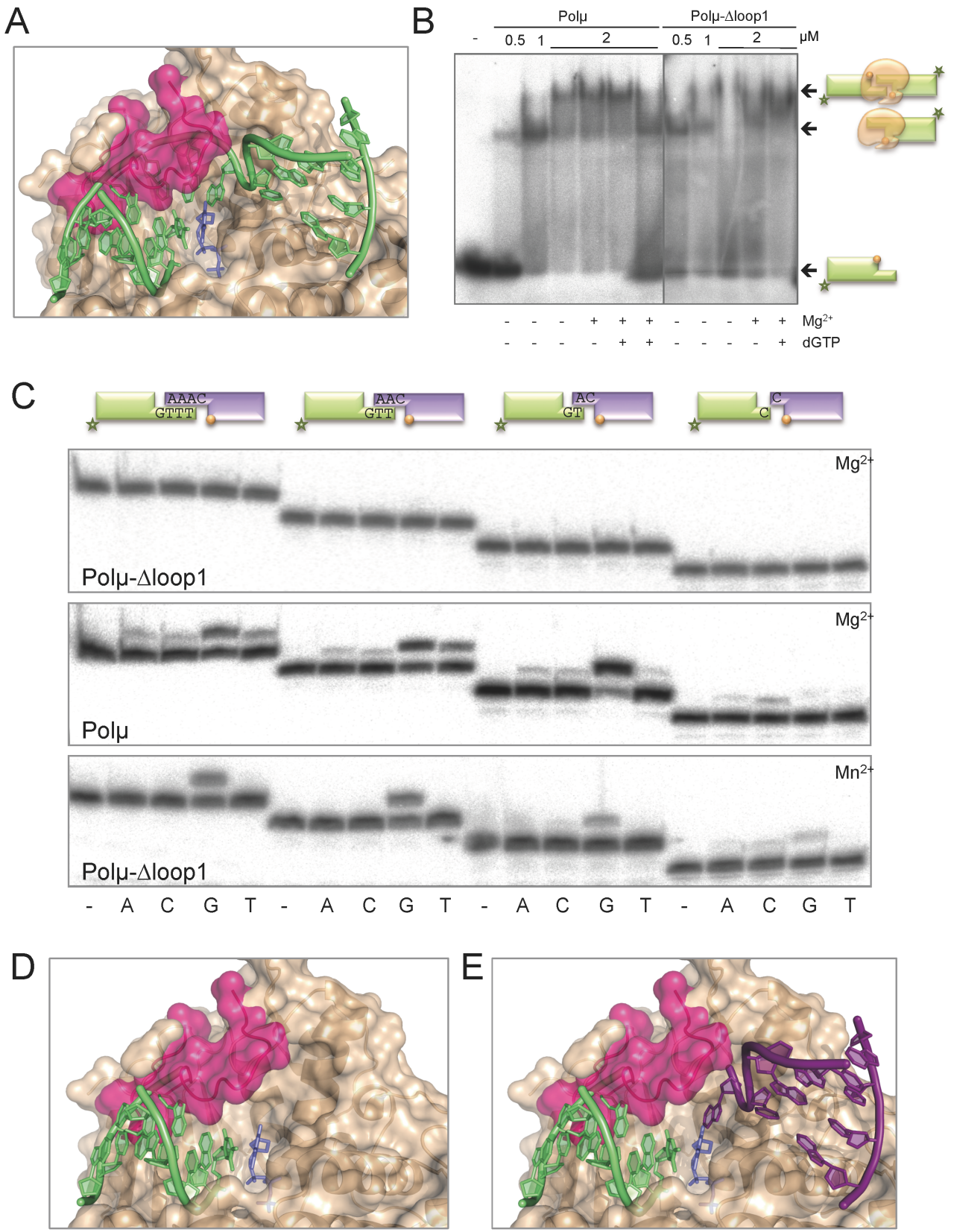



**Figure 45. Loop 1 dictates the preferred connection of two DNA ends.**
A) Superimposition of Loop1 from the crystal of the TdT apoenzyme (PDB ID: 1JMS), shown in dark pink cartoon and semi-transparenct surface, on the Pol*μ* ternary complex structure (PDB ID:2IHM), shown as a wheat-colored surface. The DNA substrate is shown in green cartoon, the incoming dNTP is shown in blue sticks. B) EMSA was performed for the wild-type and mutant Pol*μ* at the indicated amounts (0.5, 1, 2 *μ*M) using a 3'-ptrotruding substrate containing the oligonucleotides 1TG and 1D-NHEJ. When indicated, 2.5 mM $MgCl_2$ and/or 10 *μ*M dGTP were added. After electrophoresis, gel was dried and the labeled fragments were detected by autoradiography. C) NHEJ reactions were performed with 200 nM of the indicated proteins and using four sets of substrates: the labeled substrates were formed by hybridization of 1TTTG, 1TTG, 1TG or 1C with 1D-NHEJ, and the cold substrates, by hybridization of either 1AAAC, 1AAC, 1AC or 1C with 1D-NHEJ. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, each of the four ddNTPs (10 *μ*M) were added in the presence of 2.5 mM $MgCl_2$ or 1 mM $MnCl_2$. D) Superimposition of Loop1 from the crystal of the TdT apoenzyme (PDB ID: 1JMS), shown in dark pink cartoon and semi-transparenct surface, on the Pol*μ* ternary complex structure (PDB ID:2IHM), shown as a wheat-colored surface. The DNA substrate has been modified to show a 1 nt 3'-protruding T/P structure (green cartoon) and the incoming dNTP is shown in blue sticks. E) Superimposition of Loop1 from the crystal of the TdT apoenzyme (PDB ID: 1JMS), shown in dark pink cartoon and semi-transparenct surface, on the Pol*μ* ternary complex structure (PDB ID:2IHM), shown as a wheat-colored surface. The DNA substrate has been modified to show two 1 nt 3'-protruding structures (T/P: green cartoon, T/D: purple cartoon) and the incoming dNTP is shown in blue sticks.

(Davis et al., 2008); the rationale for preferring the correct nucleotide (ddATP) is that the rate limiting step of the terminal transferase reaction, i.e. the movement of the protruding primer from an inactive to an active conformation (this model will be explained in detail in the next section), is facilitated by the approaching end, which provides a templating base "in trans". If this templating base fails to be located at a proper register, the ability of Pol*μ* to catalyze template-independent additions will allow joining (as represented in Fig. 44B, scheme 3), as Ligase IV can accept a mismatched 3'-end for ligation. These results indicate that a recessive 5'-P is key for the precise joining of incompatible ends. Interestingly, and in agreement with recent findings supporting that 1 nucleotide is the optimal 3´ overhang size for the NHEJ capacity of Pol*μ* on incompatible ends (Davis et al., 2008), the effect of the 5'-P is not detected in longer overhangs (Fig. 44C). When using longer 3'-protruding incompatible substrates, the enzyme only inserts untemplated nucleotides as is unable to accommodate the DNA ends.

**2. Loop 1: effect on NHEJ and implications on the selection of the gap-length and separation.**

Loop 1 is a specific subdomain in Pol*μ*, shared with TdT, which is flexible and thus can adopt multiple conformations, and probably acts as a pseudo-template when a proper template DNA strand is not available for instructing polymerization. Moreover, due to its flexible nature, this structure was not visible in the crystal structure obtained for murine Pol*μ* (Moon et al., 2007). TdT Loop 1, however, is fixed in a position in which it blocks the accommodation of a templating strand, probably being the main reason why TdT is a completely creative polymerase (Delarue et al., 2002; Juárez et al., 2006). This domain could be crystallized in TdT due to its immobility, and in figure 45A we show a superimposition of the conformation adopted by Loop 1 of TdT in the apoenzyme (PDB ID: 1JMS) modeled on the ternary structure of Pol*μ* with gapped DNA and incoming nucleotide (PDB ID: 2IHM, wheat). In agreement with the "pseudo-template" hypothesis, Loop 1 is overlapping with the template strand, a position that Pol*μ* Loop 1 would adopt in the cases in which the substrate lacks a template strand or contains a discontinuous template. Accordingly, Loop 1 has been implicated in the terminal transferase activity of the enzyme (Juarez et al., 2006), and also has been shown to play a role during NHEJ of incompatible ends assisted by accessory factors (Nick McElhinny et al., 2005). To corroborate the importance of this particular structure for the bridging activity inherent to Pol*μ* we tested a deletion mutant (Pol*μ*-Δloop1; Juarez et al., 2006) which lacks amino-acids 369 to 385 (corresponding to the Loop 1) in our NHEJ assays in the absence of accessory factors. It has been previously shown that binding of 5'-P containing gapped substrates by Pol*μ*-Δloop1 is even higher than that displayed by the wild-type enzyme, suggesting that Loop 1 is dispensable and even detrimental for binding to template-containing substrates perhaps *via* steric hindrance (Juarez et al., 2006). As it has been already shown, wild-type Pol*μ* can form two different complexes when binding a 5'-P-containing 3'-protruding substrate. The first shifted band corresponds to the binding to one DNA molecule as a T/D substrate, while the second band represents the synapsis of two DNA molecules (shown again in Fig. 45B, left panel, to allow an easier comparison with the mutant enzyme). Binding to a single labelled molecule, representing the downstream side of the break, was also observed with mutant Pol*μ*-Δloop1, even more strongly than wild type Pol*μ* at low concentrations (Fig. 45B, right panel). This observation suggests that Loop 1 interferes the binding to a 3´overhang (as part of a discontinuous template strand), mimicking the results previously obtained with gapped DNA molecules (Juarez et al., 2006). However, the synapsis could not be stably formed in the absence of Loop 1, not even in the presence of metal and incoming nucleotide (Fig. 45B, right panel). Moreover, the activity of this mutant was undetectable in NHEJ assays, both with non-compatible and compatible ends, even using substrates that can form two and three base pairs of complementarity (Fig. 45C, top panel), when compared with the wild-type Pol*μ* (Fig. 45C, middle panel). Having shown that Loop 1 is dispensable for binding the 5'-P-containing DNA end, the defective NHEJ reaction observed can only be explained if Loop 1 plays a role in the correct juxtaposition of the incoming end acting as primer. Surprisingly, NHEJ by this mutant was partially recovered when using manganese as activating metal ion. However, we

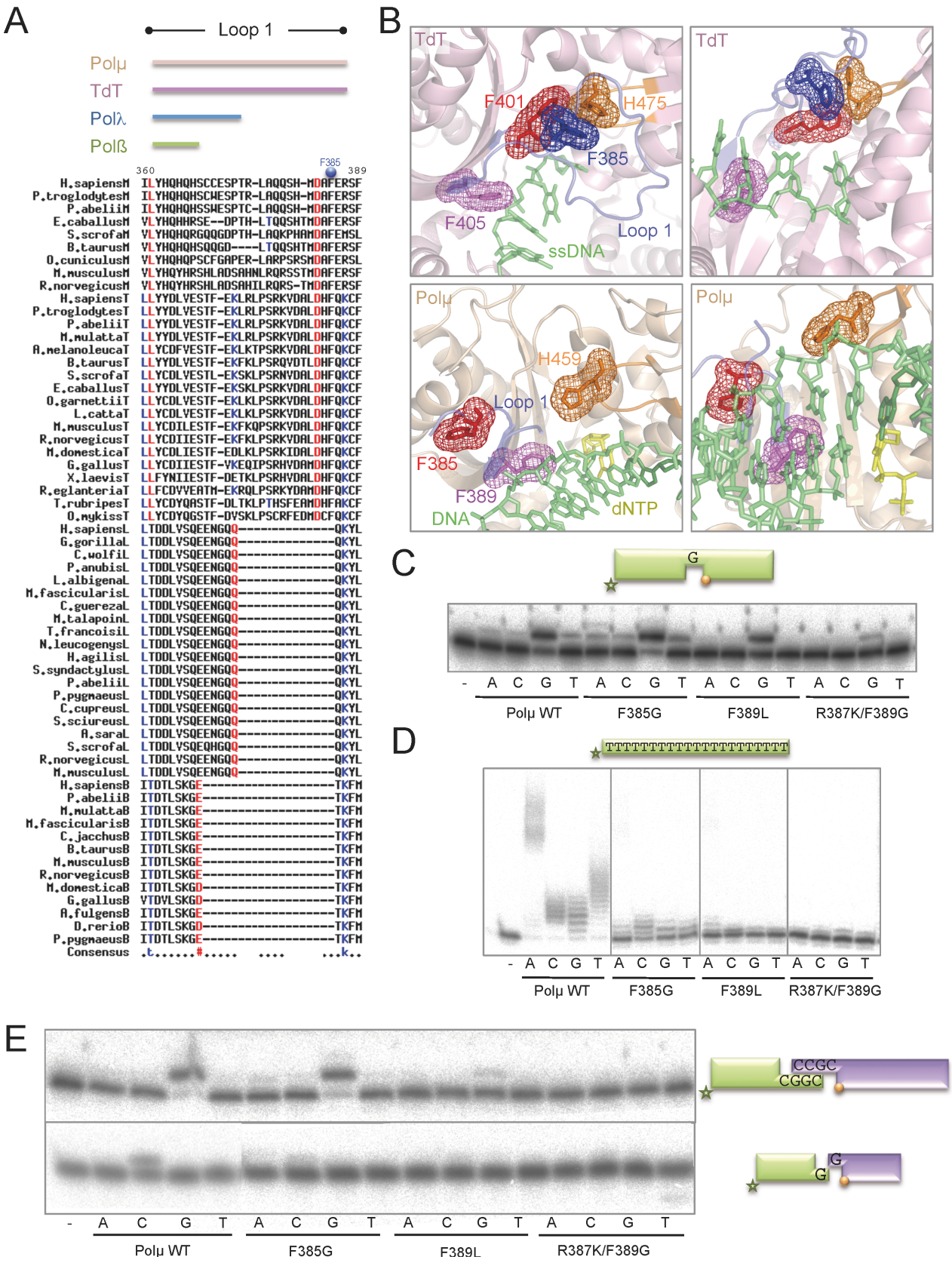



**Figure 46. The "hinge" regions of Loop 1 are implicated in the flexibility of the motif and in the efficiency of the NHEJ reaction.** A) Multiple sequence alignment showing the zone corresponding to Loop 1 in Pol*μ* (M), TdT (T), Polλ (L) and Polß (B) from different species. Numbers indicate the residue in the human Pol*μ* sequence. Residue Phe$^{385}$ is indicated with a blue dot. B) Cartoon representations of the structures of TdT bound to ssDNA (1KDH, light pink) and the Pol*μ* ternary complex (2IHM, wheat), showing the Loop 1 in blue cartoon and selected residues in sticks and mesh. C) Gap-filling reactions were peformed as described in Materials and Methods with the indicated proteins (25 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. When indicated, dNTPs were added separately at 10 nM in the presence of 2.5 mM $MgCl_2$. D) Terminal transferase activity assay with the indicated proteins (600 nM) using a homopolymeric substrate (polydT) and each of the four dNTPs (100 *μ*M). Reactions were performed for 30 min at 37ºC. E) NHEJ reactions were performed with 200 nM of the indicated proteins and using four sets of substrates: the labeled substrates were formed by hybridization of 1G with 1D-NHEJ or D3-C with D1, and the cold substrates, by hybridization of either 2G with 2D-NHEJ or D4-C with D2. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, each of the four ddNTPs (10 *μ*M) were added in the presence of 2.5 mM $MgCl_2$.

realized that while wild type Pol*μ* strikingly prefers the shortest 3' overhangs (1 nt) over longer ones (2, 3 and 4 nts) (Fig. 45C, middle panel), despite the lower levels of complementarity, the substrate preference for the mutant lacking Loop 1 has changed, achieving better NHEJ on longer protruding substrates (Fig. 45C, lower panel), as would be expected only due to the help of complementary base pairing. Therefore, Loop 1 appears to be conferring the unique Pol*μ* property of bridging ends with very short overhangs. This observation also implies that a minimal distance between the two gaps formed during end-joining, allows the optimal location, and thus performance, of Loop 1. This proposal is supported by the available crystal structures of TdT and Pol*μ*. Figure 45D&E show two different versions of the superimposition described above: in part D, the DNA substrate has been modified to show a 1 nt 3'-protruding substrate bound as a template/primer, a scenario in which Loop 1 would help to maintain the orientation of the primer strand to achieve the untemplated addition of an incoming nucleotide, shown in blue; in part E, the modification of the DNA substrate included only elimination of template bases A6 and T7 in order to mimick a non-compatible NHEJ substrate formed by a T/D molecule with a 3' overhang of 1 nucleotide (purple) and a primer molecule with a 2 nt 3' protrusion (green). As we show here, Loop 1 would fit in the space corresponding to -1 and -2 positions of a continuous template strand, and therefore, it could be conveniently positioned to help preferential bridging of NHEJ substrates with very short overhangs.

**3. Relationship between TdT activity and NHEJ efficiency: single mutations in Loop 1 affecting its structure/function.**

We have already shown the importance of Loop 1 both in terminal transferase activity and in NHEJ reactions. To perform a more in depth study of the function of this motif, we made point mutations in several residues possibly involved in its interaction with the substrates. Guided by the conservation level of different residues in the protein multi-alignments shown in figure 46A, and by comparison of the available crystal structures (Fig. 46B) of the murine TdT (PDB IDS: 1KDH, 1JMS) and Pol*μ* (PDB ID: 2IHM, monomers A and B), we decided to mutate three Pol*μ* residues included in or near Loop 1: $Phe^{385}$, $Arg^{387}$ and $Phe^{389}$.

In TdT (Fig. 46B, top panels: left panel for top view, right panel for frontal view), $Phe^{401}$ (corresponding to $Phe^{385}$ in Pol*μ*), located in the border region of Loop 1, is interacting with another highly conserved phenylalanine in the middle of this motif in TdT ($Phe^{385}$), which seems to be involved in maintaining the fixed position of Loop 1 via a strong stacking interaction between its aromatic ring and $His^{475}$ ($His^{459}$ in Pol*μ*) at the thumb subdomain. Recent data has shown that mutation of $Phe^{401}$ to alanine in TdT strongly reduced the terminal transferase activity of the polymerase and, strikingly, allows templated addition of nucleotides, turning a completely template independent enzyme into a DNA-instructed DNA polymerase (Romain et al., 2009). Our explanation to these results is that by mutating this residue the network of interactions needed to maintain a fixed orientation of TdT Loop 1 is abolished, and Loop 1 is now endowed with a greater degree of flexibility, as in Pol*μ*, thus allowing TdT to accept a template strand. Despite the sequence conservation of this phenylalanine between TdT ($Phe^{401}$) and Pol*μ* ($Phe^{385}$), there is no functional conservation, since the second phenylalanine involved is not present in Pol*μ* (Fig. 46A&B, bottom panels: left for top view, right for fontal view). We decided to mutate $Phe^{385}$ of Pol*μ* to glycine in order to stablish its role in the terminal transferase and NHEJ activity of Pol*μ*. As a control, we firstly confirmed that this residue is not implicated in gap-filling activity (Fig. 46C). Terminal transferase activity of mutant F385G, on the other hand, was largely abolished when compared to wild type Pol*μ* (Fig. 46D), thus confirming that this residue has a specific role in the catalytic cycle only when a template strand is not available. When we tested this mutation on NHEJ assays the activity of the enzyme was lower than that of the wild type Pol*μ*, but only on non-compatible ends (Fig. 46E, bottom panel), as expected since these DNA substrates contain discontinuities not only in the primer strand, but also in the template strand in a position that Loop 1 might help to coordinate.

$Phe^{389}$ is again conserved among Pol*μ*s and TdTs ($Phe^{405}$) of different species (Fig. 46A), and in both cases it is involved in maintaining the structure of this part of the protein from its location again in the bordering region of Loop 1, thus probably affecting

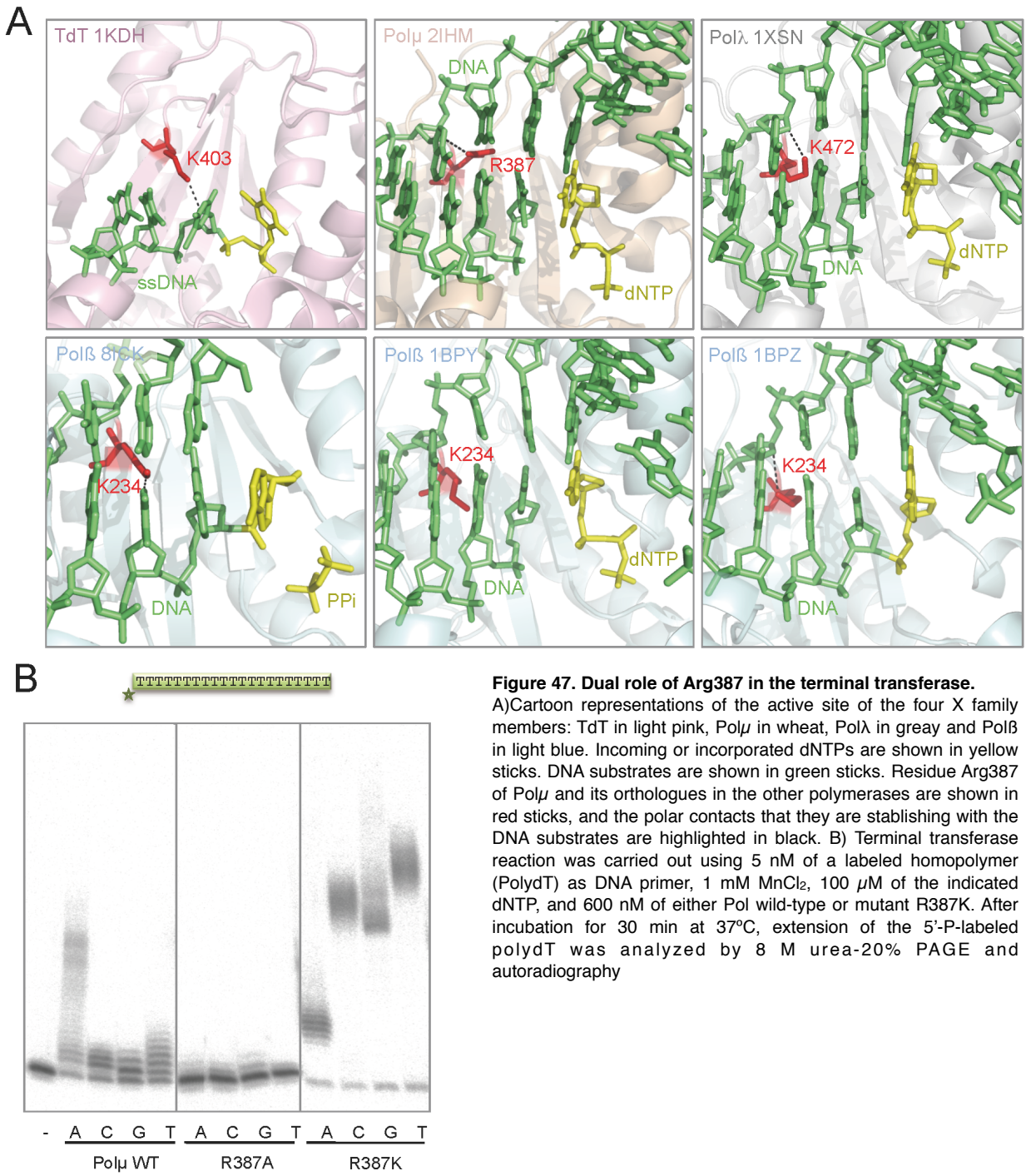



**Figure 47. Dual role of Arg387 in the terminal transferase.** A)Cartoon representations of the active site of the four X family members: TdT in light pink, Pol*μ* in wheat, Polλ in greay and Polß in light blue. Incoming or incorporated dNTPs are shown in yellow sticks. DNA substrates are shown in green sticks. Residue Arg387 of Pol*μ* and its orthologues in the other polymerases are shown in red sticks, and the polar contacts that they are stablishing with the DNA substrates are highlighted in black. B) Terminal transferase reaction was carried out using 5 nM of a labeled homopolymer (PolydT) as DNA primer, 1 mM $MnCl_2$, 100 *μ*M of the indicated dNTP, and 600 nM of either Pol wild-type or mutant R387K. After incubation for 30 min at 37ºC, extension of the 5'-P-labeled polydT was analyzed by 8 M urea-20% PAGE and autoradiography

the shape and orientation of this motif (Fig. 46B). Mutation of this residue to alanine in TdT abolishes terminal transferase activity and allows templated insertion of only one nucleotide on a template/primer substrate (Romain et al., 2009). We decided to mutate the equivalent Pol*μ* residue ($Phe^{389}$) to leucine (the aminoacid present in Polλ) to demonstrate its importance during the catalytic cycle of Pol*μ*. We also made the double mutation F389G/R387K, that could be expected to have a boosted terminal transferase activity (Andrade et al., 2009). Both mutants were tested for gap-filling activity and, as showed in figure 46C, none of them was strongly affected when compared to the wild type Pol*μ*. On the other hand, the expected implication of $Phe^{389}$ in the ability of Pol*μ* to catalyze untemplated nucleotide additions was confirmed by testing the terminal transferase activity of these mutants: it was completely abolished (Fig. 46D), also in the case of the double mutant, which is specially noticeable since this enzyme also contains the R387K mutation that, alone, increases this activity by 100-fold

(Andrade et al., 2009). When NHEJ activity was assayed (Fig. 46E), the two mutants were completely negative, as expected from an affected orientation of Loop 1, and the consequent discoordination of the DNA substrates during this kind of reactions.

Another residue that could be affecting the structure/function of Loop 1 is Arg$^{387}$ (Lys$^{403}$ in TdT), conserved among Pol$\mu$'s of different species, but absent in the other three members of the family X that have a lysine in this position (see alignment in figure 46A). In the crystal structure of TdT with a ssDNA (Fig. 47A, top left) Lys$^{403}$ is interacting with the DNA substrate. This ssDNA is located in an unproductive position, with its 3'-terminus occupying the incoming nucleotide position. There is one Polß structure available (PDB ID: 8ICK; Fig. 47A, bottom right) in which the DNA substrate, despite being double stranded instead of single stranded, is occupying the same unproductive position as that described in the TdT crystal (Fig. 47A, top left): in this post-catalytic structure, the 3'-*terminus* of the primer is still located at the position of the incoming nucleotide, since translocation has not taken place. In the Polß structure, Lys$^{234}$ (corresponding to Arg$^{387}$ in Pol$\mu$ and Lys$^{403}$ in TdT) is also contacting the primer strand (Fig. 47A, bottom right). On the other hand, in the structure of Pol$\mu$ with a gapped DNA and incoming nucleotide (PDB ID: 2IHM), Arg$^{387}$ is interacting with the -3 position of the template strand, the same as Lys$^{472}$ in the Polλ ternary structure (PDB ID: 1XSN; Fig. 47A, top middle &

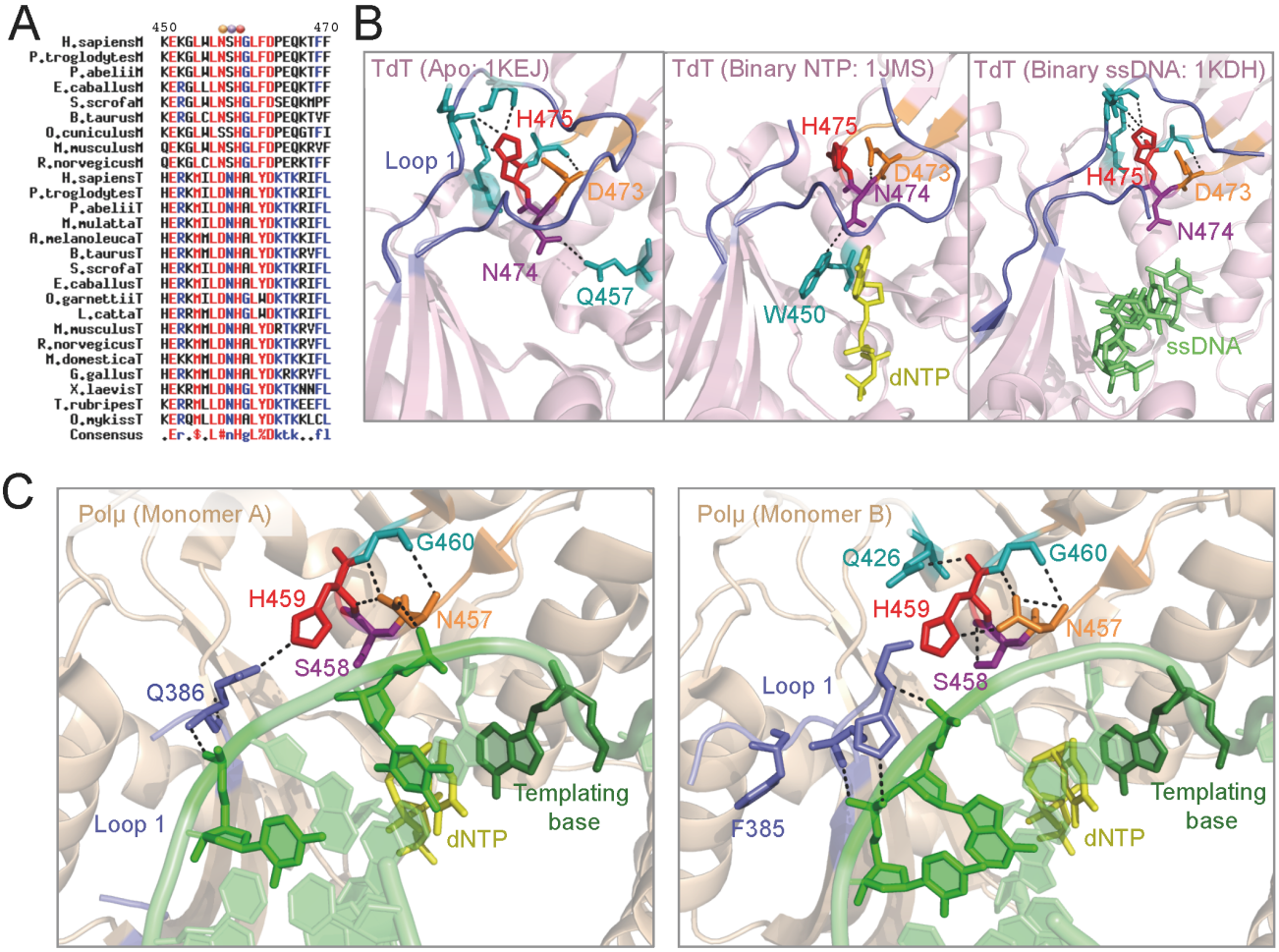



**Figure 48. The thumb "mini-loop" (NSH motif) of Pol$\mu$ is implicated in the correct positioning of Loop 1 and template strand.** A) Multiple sequence alignment showing the NSH motif region in Pol$\mu$ (M) and TdT (T) from different species. Red color indicates invariant residues, blue color indicates conservative changes. The three mutated residues are indicated with dots. B) Cartoon representatin of the three available TdT structures showing Loop 1 in blue cartoon and the thumb "mini-loop" in orange cartoon. The mutated residues are shown in orange, red and purple sticks, while the residues included in their network of contacts are shown in teal-colored sticks. Incoming dNTP is shown in yellow sticks, DNA substrate is shown in green sticks. C) Cartoon representatin of the two monomers included in the Pol$\mu$ crystal structure showing Loop 1 in blue cartoon and the thumb "mini-loop" in orange cartoon. The mutated residues are shown in orange, red and purple sticks, while the residues included in their network of contacts are shown in teal-colored sticks. Incoming dNTP is shown in yellow sticks, DNA substrate is shown in green sticks.

right, respectively). When Polß was complexed with a gapped DNA, the corresponding lysine shows different conformations depending on the step of the catalytic cycle captured: in the pre-catalytic complex of Polß with a gapped substrate and incoming nucleotide (PDB ID: 1BPY), Lys$^{234}$ is oriented towards the primer strand (Fig. 47A, bottom middle panel), while in the post-catalytic, but pre-translocation, step of the same reaction (structure of Polß with nicked DNA, PDB ID: 1BPZ), Lys$^{234}$ is contacting the template strand, also at the -3 position as Pol*μ* and Polλ (Fig. 47A, bottom right). In order to study the possible role of the dual interactions stablished by this residue during the catalytic cycle, we decided to use two mutants already existing in the lab, in which this residue was changed either to alanine, to eliminate the reactivity of the residue, or to lysine, in order to mimick the sequence conserved in Polß, Polλ and TdT. As expected, we showed that these mutants were not affected in gap-filling activity, behaving in this context exactly like wild type Pol*μ* but, strikingly, they showed very different phenotypes when tested for terminal transferase activity (Andrade et al., 2007). Figure 47B shows that mutant R387A has almost null activity for untemplated additions, while mutant R387K displayed a strong increase in the terminal transferase activity, of even 100-fold with certain substrates. The differential phenotype of these mutants and the implications of this residue for terminal transferase and NHEJ activities of Pol*μ* will be described in more detail later.

**4. The thumb mini-loop: affecting flexibility of Loop 1.**

By analyzing the available structures of TdT (PDB IDS: 1JMS, 1KEJ and 1KDH) we found a second loop, located in the thumb subdomain, that is stablishing interactions with Loop 1. This thumb mini-loop contains several conserved residues responsible for these interactions, among them the invariant Asp$^{473}$, Asn$^{474}$ and His$^{475}$ (Fig. 48B, DNHmotif). This thumb mini-loop is also present in Pol*μ*, but the sequence is not strictly conserved (Fig. 48A, NSH motif). The only invariant residues are the histidine, His$^{459}$ in Pol*μ*, and a leucine, Leu$^{461}$ in Pol*μ*. A small side chain (Ala or Gly) is present between these two. The other two residues present in the TdT sequence, Asp$^{473}$ and Asn$^{474}$, are an asparagine (Asn$^{457}$) and a serine (Ser$^{458}$), respectively, in Pol*μ* (Fig. 48C).

In TdT, Asp$^{473}$ seems to be involved in maintaining the general conformation of this mini-loop through interactions with other residues from the motif (Ala$^{476}$, Asn$^{474}$). Asn$^{474}$ changes very slightly its orientation in the three TdT crystals available, but this minor movement is enough to stablish different interactions in each case: in the Apo structure this residue is interacting with Glu$^{457}$, a residue that is only at contact distance when no substrate is present (Fig. 48B, left panel); in the NTP-bound structure Asn$^{474}$ is contacting Trp$^{450}$ probably due to its stacking against the incoming nucleotide that slightly affects its position (Fig. 48B, central panel); finally, in the enzyme-ssDNA co-crystal, this asparagine is not making any contacts, since none of its partners is available for interaction (Fig. 48B, right panel). His$^{475}$ is stablishing a strong network of direct interactions with Loop 1, through residues Glu$^{386}$ and Lys$^{387}$ or Lys$^{389}$, depending on the crystal structure studied. It is also making direct interactions with residue Arg$^{442}$. There is one TdT crystal in which part of Loop 1 (corresponding to the residues involved in these interactions) has not been crystallized: the binary complex with incoming nucleotide (Fig. 48B, central panel), in which His$^{475}$ has rotated and now this stabilizing network is disrupted.

All these observations led us to study the possible role of this thumb mini-loop in Pol*μ*, that could also have a Loop 1-stabilizing function, not as constitutively as in TdT, but perhaps specifically during NHEJ. The available Pol*μ* structure in complex with a gapped-substrate and incoming nucleotide (Fig. 48C), includes two monomers in the unit cell, and the two of them show different states and interactions of residue His$^{459}$ (His$^{475}$ in TdT): in one of the monomers is interacting with Glu$^{386}$ from Loop 1, thus mimicking the function of this residue in TdT, but in the other it is contacting Asn$^{457}$ (Asp$^{473}$ in TdT). This strong and very close interaction stablished with the histidine seems to be capturing it and preventing its interaction with its otherwise partner in Loop 1, probably allowing this motif to be flexible and move away to allow binding of a template strand. When the template strand is not present (terminal transferase) or is discontinuous (NHEJ), His$^{459}$ might help stabilizing Loop 1 through the described interaction with Glu$^{386}$, in a position

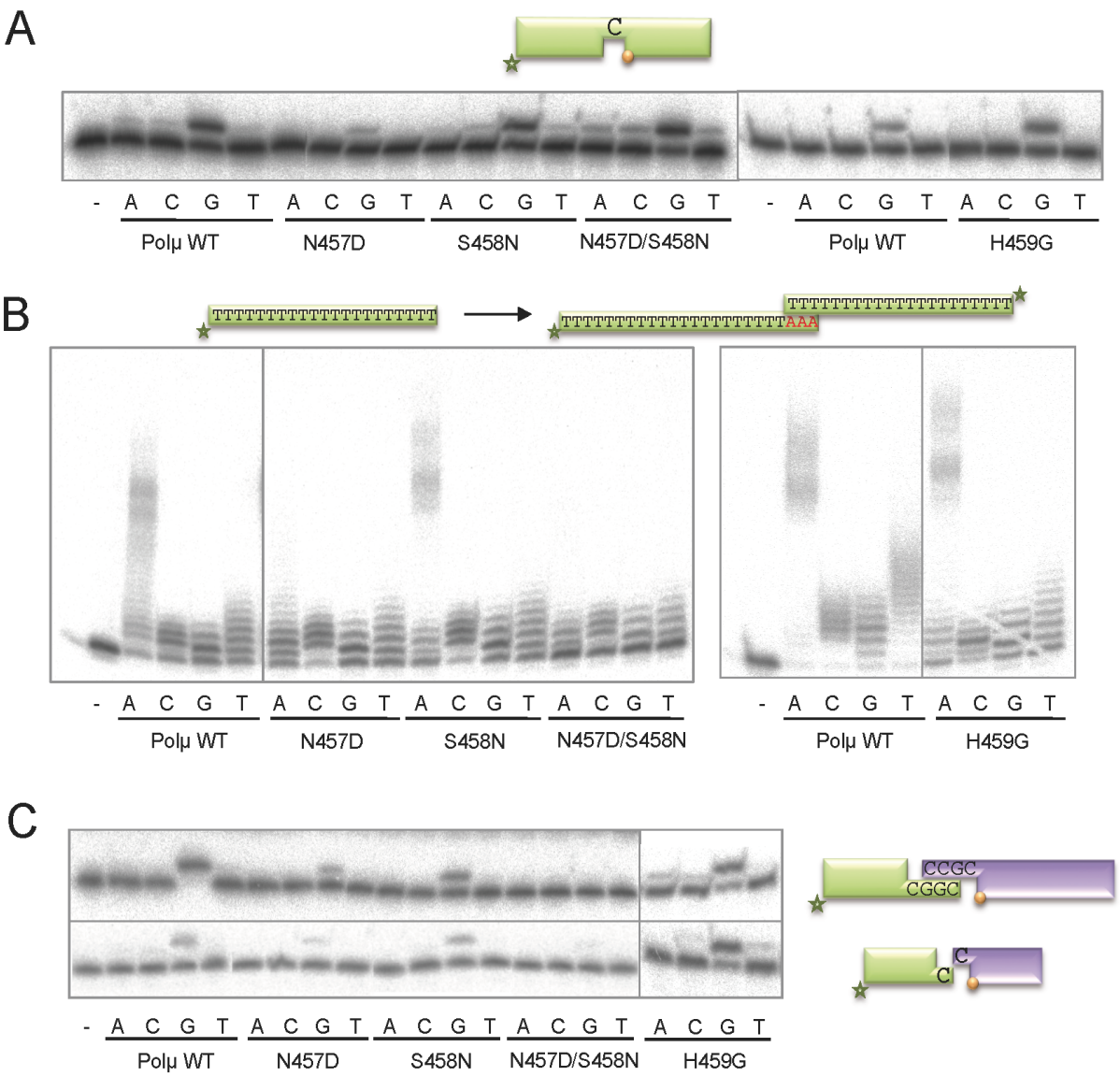



**Figure 49. Mutations in the thumb "mini-loop" (NSH motif) of Pol*μ* specifically affect terminal transferase and NHEJ.**
A) Gap-filling reactions were performed as described in Materials and Methods with the indicated proteins (25 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. When indicated, dNTPs were added separately at 10 nM in the presence of 2.5 mM $MgCl_2$. B) Terminal transferase activity assay with the indicated proteins (600 nM) using a homopolymeric substrate (polydT) and each of the four dNTPs (100 *μ*M). Reactions were performed for 30 min at 37ºC. C) NHEJ reactions were performed with 200 nM of the indicated proteins and using four sets of substrates: the labeled substrates were formed by hybridization of 1G with 1D-NHEJ or D3-C with D1, and the cold substrates, by hybridization of either 2G with 2D-NHEJ or D4-C with D2. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, each of the four ddNTPs (10 *μ*M) were added in the presence of 2.5 mM $MgCl_2$.

more similar to that observed in the TdT crystals. Interestingly, when $His^{459}$ is interacting with $Glu^{386}$, $Asn^{457}$ directly contacts the template strand (Fig. 48C, monomer A). Given the possible importance of Pol*μ* $His^{459}$, as well as its "trapping" $Asn^{457}$, we decided to prepare mutants H459G to abolish the function, N457D, S458N, and the double mutation N457D/S458N in order to mimick the residues present in TdT in this area.

The behaviour of these mutants in gap-filling activity (i.e in the presence of a continuous template strand) was different in each case: mutant N457D had lower activity than the wild type on this substrate, in agreement with its observed interaction with the template strand; S458N, double mutant N457D/S458N and H459G were able to perform gap-filling as the wild type Pol*μ* (Fig. 49A). When testing terminal transferase activity, mutant S458N was the only one that displayed a similar level of activity to that of the wt enzyme with any of the four dNTPs. Strikingly, both the single and double mutant having the change N457D were exclusively affected in the addition of dA nucleotide units, having wild-type levels of terminal transferase activity with dC, dG

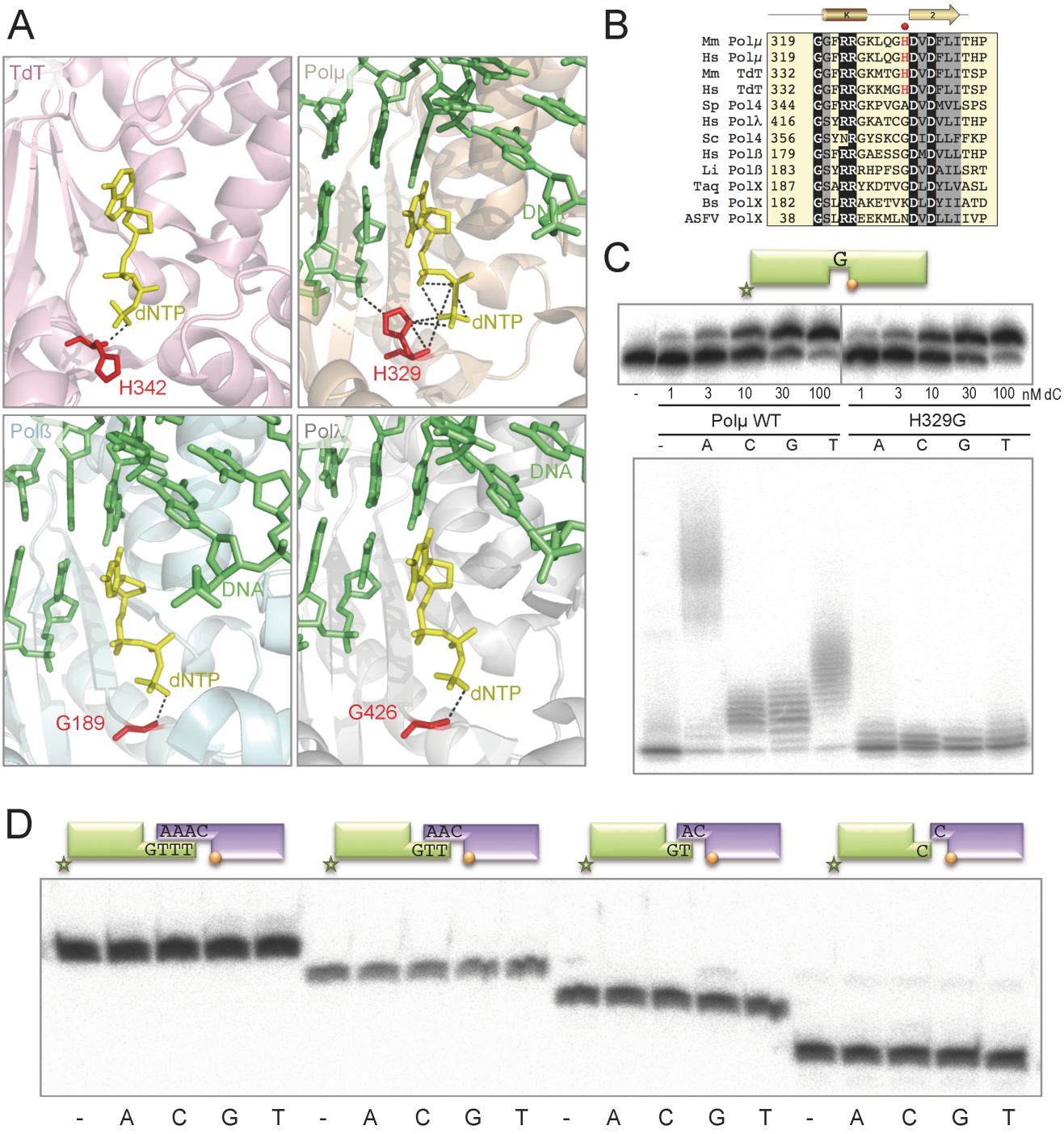



**Figure 50. His$^{329}$ is involved in both terminal transferase and NHEJ.**
A) Crystal structures of TdT complexed with dNTP (1KEJ) and Pol*μ* (2IHM), Polß (1BPZ) and Polλ (1XSN) ternary complexes. Residue His$^{329}$ in Pol*μ*, and its orthologues, are indicated in red sticks. DNA substrates are indicated in green. Incoming ddNTPs are shown in yellow. B) Amino acid sequence alignment of family X polymerases from different species (*Hs, human; Mm, mouse; Sp, Schizosaccharomyces pombe; Sc, Saccharomyces cerevisiae; Li, Leishmania infantum; Taq, Thermus aquaticus; Bs, Bacillus subtilis*; and ASFV, african swine fever virus) along the indicated regions. Residue His$^{329}$ in Pol*μ* is indicated with a red dot. Invariant residues are depicted in white over black background, and conservative residues have gray background. C) Gap-filling reactions (top panels) were performed as described in Materials and Methods with the indicated proteins (25 nM) using a gapped substrate containing the oligonucleotides SP1C, T13G and DG1-P. Complementary dNTP (dC) was added separately at the indicated concentrations in the presence of 2.5 mM $MgCl_2$. Terminal transferase activity assay (lower panel) was performed with the indicated proteins (600 nM) using a homopolymeric substrate (polydT) and each of the four dNTPs (100 *μ*M). Reactions were carried out for 30 min at 37ºC. D) NHEJ reactions were performed with 200 nM of the indicated proteins and using four sets of substrates: the labeled substrates were formed by hybridization of 1TTTG, 1TTG, 1TG or 1C with 1D-NHEJ, and the cold substrates, by hybridization of either 1AAAC, 1AAC, 1AC or 1C with 1D-NHEJ. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, each of the four ddNTPs (100 *μ*M) were added in the presence of 2.5 mM $MgCl_2$.

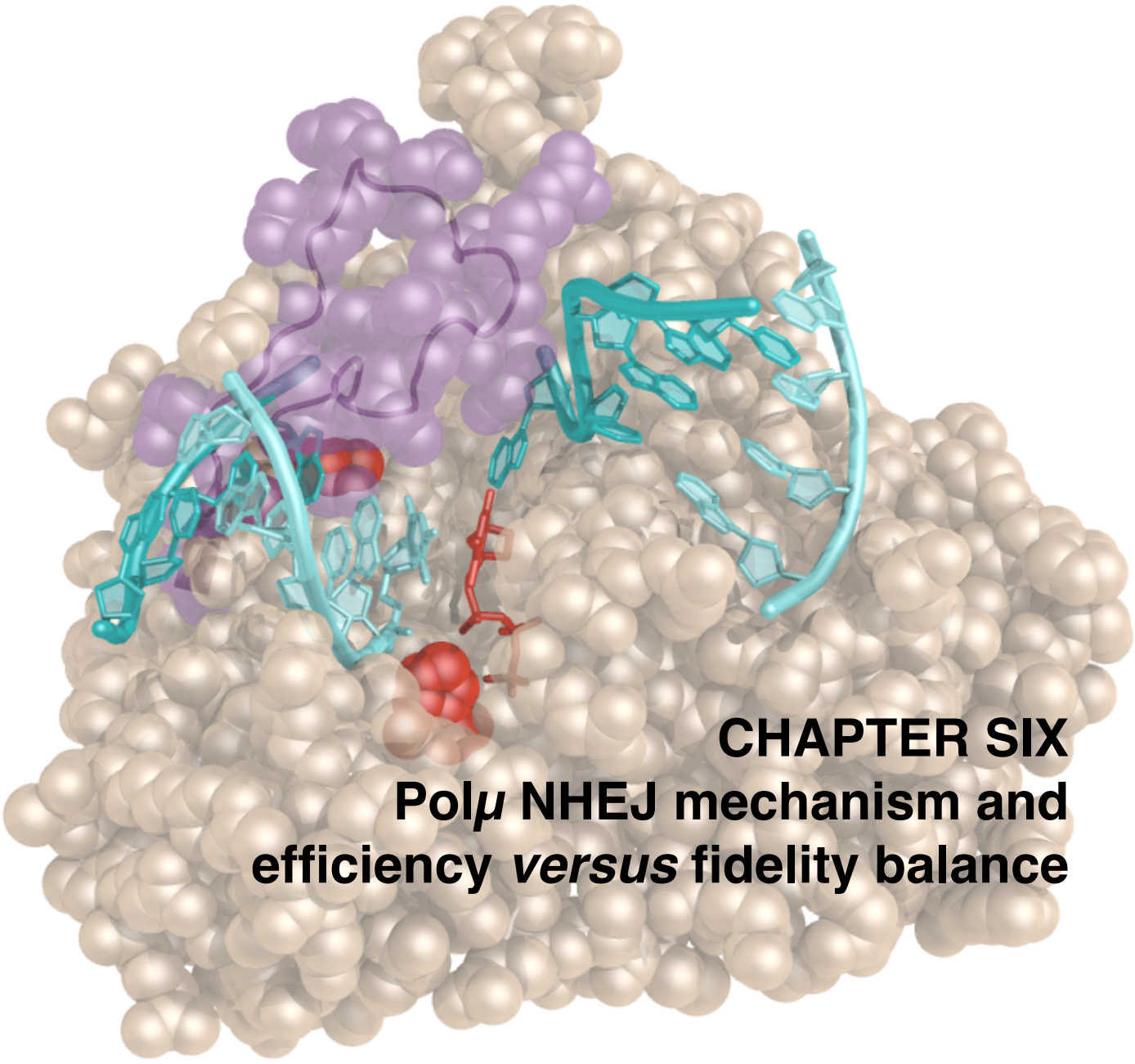
and dT (Fig. 49B). The extremely long (dA)n products produced by the wild-type Pol*μ* have been interpreted mainly as the result of DNA-templated incorporation, allowed by the connection of a dA-extended PolydT (*via* terminal transferase) with another PolydT molecule, now acting as template (Juarez et al., 2006; see scheme at Fig. 49B). Therefore, it is very likely that mutation N457D is not affecting terminal transferase, but precludes the connection/synapsis step that is allowing template-directed incorporation of dATP. This implies that the DNA ligand function of Asn$^{457}$ observed in a DNA gap is crucial for synapsis of complementary ends during NHEJ. Conversely, mutant H459G maintained normal levels of dA incorporation on PolydT, while addition of the other 3 dNTPs (*via* terminal transferase) was significantly inhibited. That would be in agreement with a role of His$^{459}$ in maintaining the correct orientation of Loop 1 (through its interaction with Glu$^{386}$) required specifically for terminal transferase, but it appears not to be critical for the function of Loop 1 during synapsis of two compatible ends.

To evaluate this hypothesis, we assayed these mutants on NHEJ reactions involving either complementary or non-complementary ends. As expected, residue His$^{459}$ seems to be unnecessary since mutant H459G reached the same activity levels as the wilt Pol*μ* on every substrate tested (Fig. 49C). Mutant N457D was largely affected in NHEJ reactions involving both complementary and non-complementary ends, although it maintained an error-free outcome (dG being preferentially inserted; Fig. 49C). Interestingly, single mutant S458N and the double mutant N457D/S458N, which had a wild-type behaviour on gap-filling reactions, were completely unable to perform NHEJ of both complementary and non-complementary ends.

All these observations lead to the conclusion that this NSH motif (thumb mini-loop) in Pol*μ* is playing a role during terminal transferase additions, through stabilization of Loop 1 in the absence of a template strand (thus mimicking the function of His$^{475}$ in TdT), but specially during NHEJ reactions, probably by stablishing interactions with the template strand that improve the connection of the two ends.

**5. Residue His$^{329}$: implications for terminal transferase and NHEJ.**

Residue His$^{329}$ is conserved only in TdT (His$^{342}$) and Pol*μ*, while Polλ and Polß possess a glycine in this position (Gly$^{426}$ and Gly$^{189}$, respectively). These residues interact with the gamma phosphate of the incoming nucleotide in all members of the X family, but also interact with the 3'-terminus of the primer in Pol*μ* and TdT (Fig. 50A). Since a histidine is only shared by the polymerases with terminal transferase activity, and because of its role in binding and coordinating the substrates implicated in this reaction, it was a perfect candidate as a residue implicated in the terminal transferase activity of Pol*μ*. Previous work in the laboratory showed that indeed, when His$^{329}$ was mutated to glycine (the residue present in Polß and Polλ), terminal transferase activity of human Pol*μ* was totally abolished, without affecting templated additions during gap-filling (Andrade et al., 2009; Figure 50B&C). In agreement with our observations, a recent publication also showed the essential role of this residue for the terminal transferase activity of Pol*μ* by mutating it to alanine (Moon et al., 2007). This residue is, therefore, an example of a single aminoacid specifically involved in allowing untemplated additions, but irrelevant during templated incorporations. So, we decided to test mutant H329G in our NHEJ assays to check the importance of the terminal transferase mechanism during end-joining. As shown in figure 50D, mutant H329G was unable to correctly perform catalysis on any substrate, either compatible or incompatible ends, whereas wild-type Pol*μ* performed an efficient and templated addition on any of them under the same conditions (Fig. 45). Again, this evidence strongly suggests that the mechanism of catalysis that occurs during the NHEJ reaction involves residues and interactions very similar to those implicated in the terminal transferase reaction.



# CHAPTER SIX
# Pol*μ* NHEJ mechanism and efficiency *versus* fidelity balance

**1. The "brake" model: Arg$^{387}$ regulates the rate of terminal transferase activity in human Pol*μ*.**

Our previous data (Andrade et al., 2009) showed that neither mutant R387K, nor mutation of the same residue to alanine, displayed any difference in gap-filling activity with respect to the wild type Pol*μ*, meaning that Arg$^{387}$ does not play a critical role in template-directed synthesis. Conversely, when terminal transferase was measured on various substrates, the effect of the mutations was dramatic: mutation R387A completely abolished activity, while R387K (mimicking TdT) produced a very significant increase of this template-independent reaction in any of the substrates tested. These results suggested that this specific arginine, that is only present in Pol*μ* but not in the other PolX members, could act as a regulator, reducing the catalytic efficiency of its intrinsic terminal transferase. The specific Arg$^{387}$ in Pol*μ* somehow limits the catalytic step of the terminal transferase reaction, perhaps explaining its low activity in comparison to TdT.

The first clue of a plausible rate-limiting step for the terminal transferase derived from the crystal structure of TdT bound to ssDNA (PDB ID: 1KDH), whose 3´-terminus was unexpectedly located in a post-catalytic situation, occupying the binding site for the next incoming dNTP (Delarue et al., 2002). No crystal structure of a pre-catalytic ternary complex for the terminal transferase reaction is available, neither for TdT nor for Pol*μ*. Therefore, and combining the 3D information available for both TdT and Pol*μ*, it can be proposed that the rate-limiting step of the terminal transferase most likely involves, for both enzymes, the movement of the single-stranded (3'-protruding) primer-terminus from a "non-productive" E:DNA complex to a "productive" E:DNA:dNTP complex (see scheme at figure 51), to achieve untemplated addition of nucleotides, irrespective of either a processive or a distributive cycling of the enzyme. To support this rate-limiting step hypothesis, we evaluated the stability of the ssDNA primer in the absence or presence of metal and dNTP (that will allow formation of a productive complex), but impeding reaction by the presence of a ddNTP at the primer-*terminus*. We firstly assessed that human Pol*μ* stably interacts with ssDNA, in a perhaps non-productive complex as that observed in the crystal structure of TdT. Paradoxically, mutant R387K, although having an increased terminal transferase activity, displayed an about 4-fold lower interaction with ssDNA. Interestingly, binding of both wild type and R387K mutant Pol*μ* to ssDNA was weakened by the presence of dNTP (Fig. 52A),

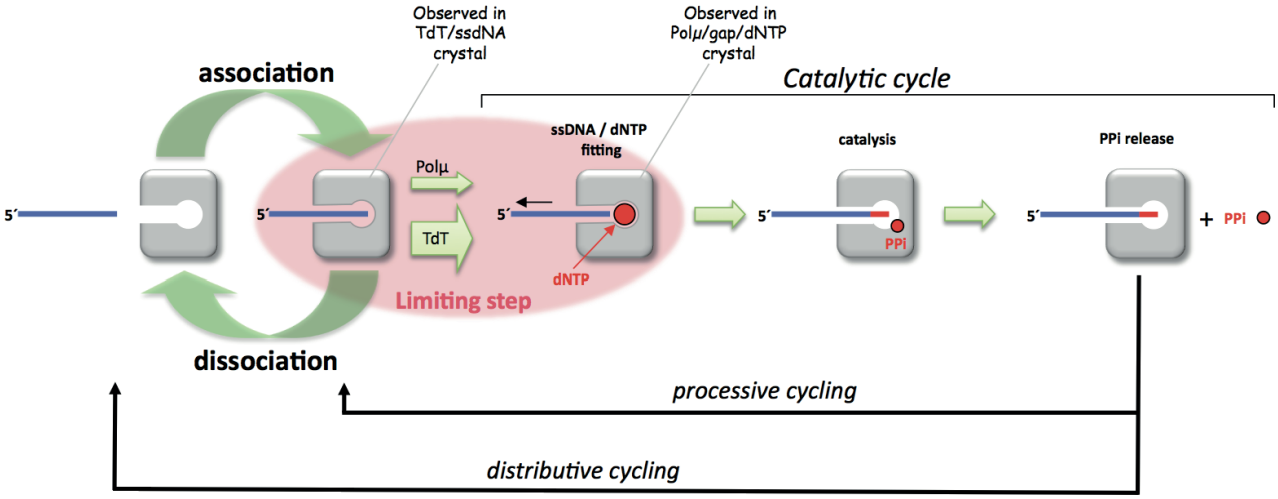



**Figure 51. Limiting step for terminal transferase addition of nucleotides.**
Untemplated terminal transferase addition of nucleotides on ssDNA requires several steps: 1) Formation of a binary complex between the enzyme and ssDNA. Upon association, that binary complex is stable but unproductive, according to TdT structural information. 2) To obtain a catalytically competent ternary complex, relocation of the primer terminus has to occur, driven by the incoming nucleotide, but in the absence of a templating base. That conformational change imposes a rate-limiting step for untemplated addition of nucleotides, that restricts terminal transferase capacity to specialized enzymes as TdT and Pol. According to our data, the different terminal transferase efficiency of these two enzymes depends on their intrinsic capacity to overcome this rate-limiting step. 3) Further steps in the catalytic cycle will include catalysis and PPi release. Reiterative synthesis, either being distributive or processive, will always face that rate-limiting step to achieve enzyme/ssDNA/dNTP fitting

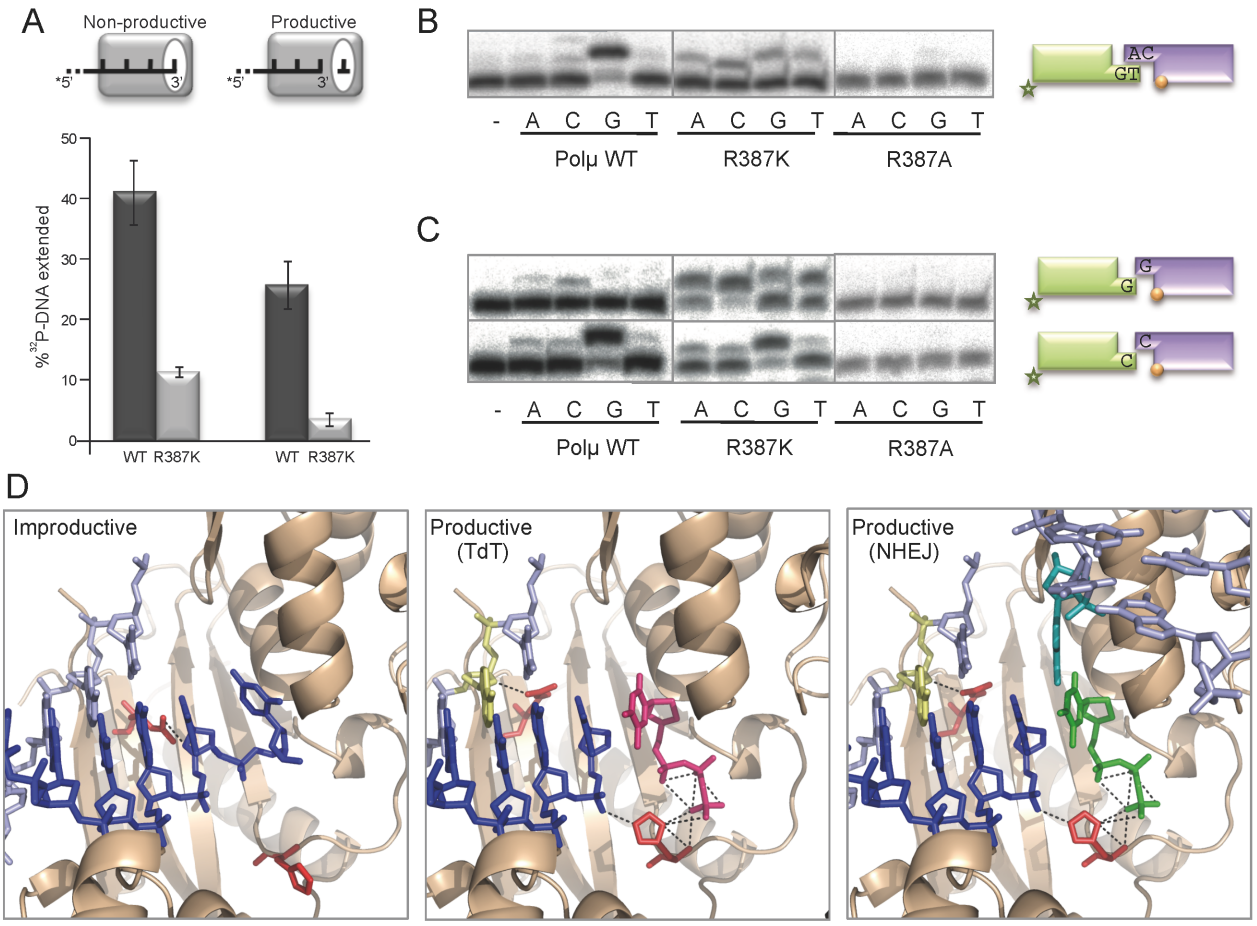



**Figure 52. Role of Arg$^{387}$ in the NHEJ mechanism**
A) Mutants at residue Arg$^{387}$ have a reduced ssDNA binding affinity. The presence of dNTP reduces ssDNA binding capacity by Pol$\mu$. B) Schemes depicting two different complexes: non-productive (nucleotide absent), indicative of a E:DNA binary complex in which the 3' end of primer is occupying the nucleotide pocket (white oval); and productive (nucleotide present), indicative of a E:DNA:dNTP ternary complex in which the primer has been relocated. Binding capacity of Pol$\mu$ wt and R387K mutant (600 nM in each case) to ssDNA/polyT, either in the absence (left) or presence (right) of dTTP, was measured by nitrocellulose filter binding assay, as described in Materials and Methods. B&C) R387K has a low fidelity during NHEJ of incompatible ends. B) Accurate NHEJ of minimally complementary ends, using a 5'-labeled 3'-protruding (GT) dsDNA substrate (5 nM), and a cold 3'-protruding (CA) dsDNA substrate (25 nM), both having a recessive 5'-P. Mutant R387K displayed an accurate behavior comparable to wild-type Pol$\mu$. C) Top panel: Inaccurate NHEJ of incompatible ends, using a 5'-labeled 3'-protruding (G) dsDNA substrate (5 nM), having a recessive 5'-P. Mutant R387K was more efficient but displayed an inaccurate behavior compared to wild-type Pol$\mu$. Lower panel: Accurate vs. inaccurate NHEJ of incompatible ends, using a 5'-labeled 3'-protruding (C) dsDNA substrate (5 nM), having a recessive 5'-P. Mutant R387K displayed an error-prone behavior compared to the accuracy of wild-type Pol$\mu$. In all cases, assays were performed in the presence of 2.5 mM MgCl2, 100 $\mu$M of the indicated ddNTP, and 200 nM of either Pol$\mu$ wt or R387K mutant. After incubation for 30 min at 30 °C, +1 extension of the 5'-P-labeled primer was analyzed by 8 M urea- 20% PAGE and autoradiography. D) Modeling the limiting step of Pol$\mu$'s terminal transferase. Left panel: Stable/non-productive step (binary complex): Pol$\mu$ (1IHM) was overimposed on the binary complex of TdT with ssDNA (1KDH). Only the partial structure of Pol$\mu$ (wheat color) is shown for clarity. A 1nt 3'-protruding substrate (derived from the gapped substrate present in the Pol$\mu$ crystal) was modeled (as in 1KDH) to reproduce the initial situation of a binary complex in which the primer (dark blue) occupies the NTP binding pocket. Residue His$^{329}$ (as His$^{342}$ in TdT) is in "standby," and Arg$^{387}$ (modeled from Lys$^{403}$ in TdT) strongly interacts with the primer strand. Middle panel: Productive step for the terminal transferase (ternary complex in the absence of template): the primer (dark blue) has been relocated, and any incoming dNTP (magenta) sits in place with the assistance of His329 (rotated 180º). Arg$^{387}$ stabilizes the new primer location by interacting with the template strand (nucleotide in yellow). Righ panel: Productive step during NHEJ (ternary complex with a template provided in trans): a NHEJ reaction implying two 3'-protruding and incompatible ends (modeled from the gapped substrate present in the Pol$\mu$ crystal) is shown. Primer relocation and Arg$^{387}$ repositioning is facilitated by the template provided in trans (nucleotide in cyan) that allows selection of a complementary nucleotide (depicted in green).

supporting the idea that there is a necessary movement of the primer (1 nt backwards) in order to form a productive complex, in which the dNTP has to compete and finally displace the 3´-terminus to occupy their corresponding binding sites, necessary for catalysis. Congruently, the intrinsic weaker binding to ssDNA of mutant R387K would facilitate primer movement associated to the rate limiting step, explaining the paradox of its high terminal transferase activity.

As expected from its enhanced terminal transferase activity and facilitated rate-limiting step, mutant R387K showed quantitative but also qualitative differences when compared to the wild type enzyme on NHEJ assays. To show this, we used a different set of 3'-protruding substrates, either with compatible (Fig. 52B) or incompatible (Fig. 52C) 3'-protruding ends. On compatible ends, both wild type Pol$\mu$ and R387K preferentially inserted the correct (templated) nucleotide (ddGTP), indicating that NHEJ of minimally compatible ends had occurred.

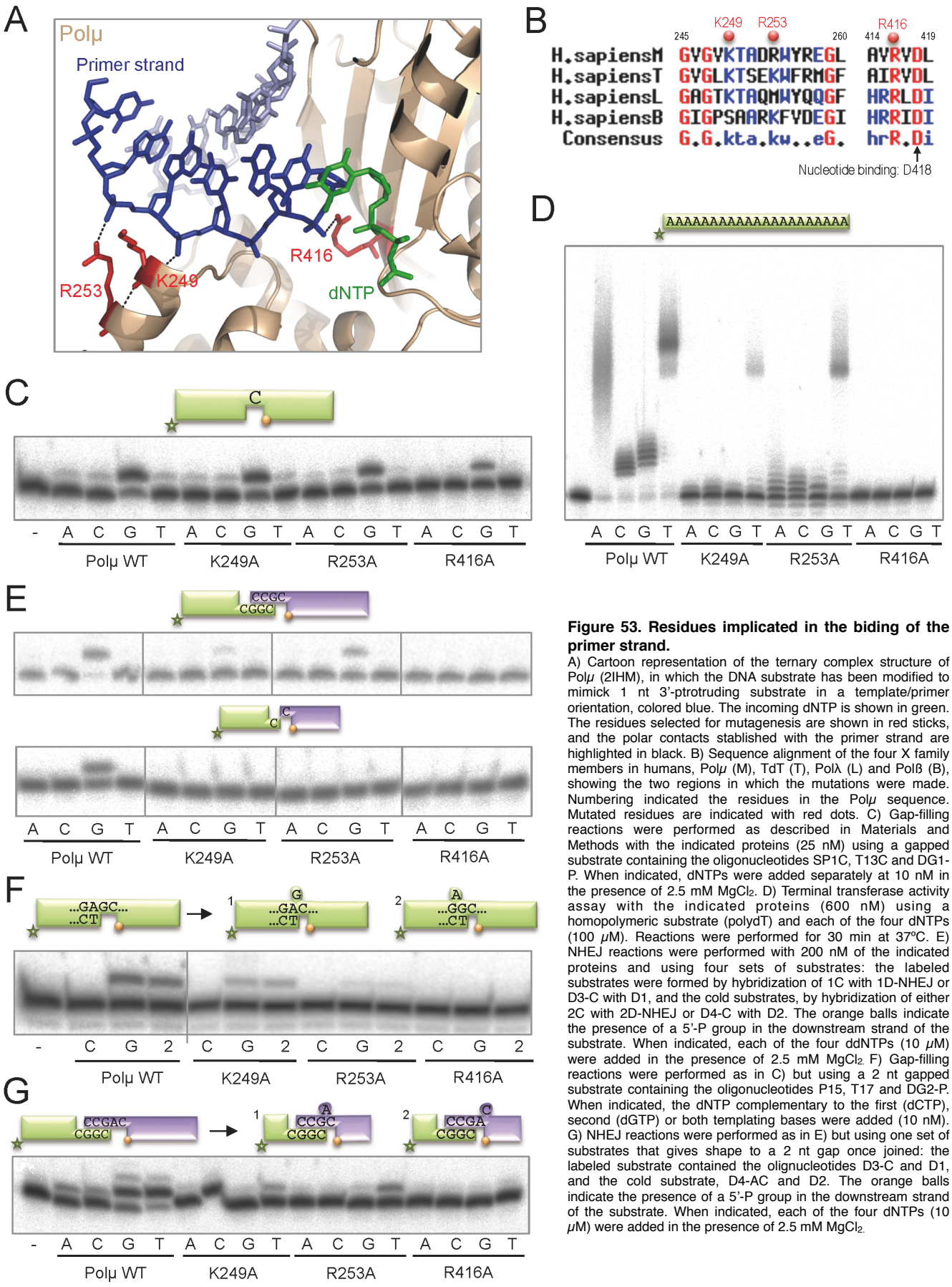



**Figure 53. Residues implicated in the biding of the primer strand.**
A) Cartoon representation of the ternary complex structure of Polμ (2IHM), in which the DNA substrate has been modified to mimick 1 nt 3'-ptrotruding substrate in a template/primer orientation, colored blue. The incoming dNTP is shown in green. The residues selected for mutagenesis are shown in red sticks, and the polar contacts stablished with the primer strand are highlighted in black. B) Sequence alignment of the four X family members in humans, Polμ (M), TdT (T), Polλ (L) and Polß (B), showing the two regions in which the mutations were made. Numbering indicated the residues in the Polμ sequence. Mutated residues are indicated with red dots. C) Gap-filling reactions were performed as described in Materials and Methods with the indicated proteins (25 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. When indicated, dNTPs were added separately at 10 nM in the presence of 2.5 mM $MgCl_2$. D) Terminal transferase activity assay with the indicated proteins (600 nM) using a homopolymeric substrate (polydT) and each of the four dNTPs (100 μM). Reactions were performed for 30 min at 37ºC. E) NHEJ reactions were performed with 200 nM of the indicated proteins and using four sets of substrates: the labeled substrates were formed by hybridization of 1C with 1D-NHEJ or D3-C with D1, and the cold substrates, by hybridization of either 2C with 2D-NHEJ or D4-C with D2. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, each of the four ddNTPs (10 μM) were added in the presence of 2.5 mM $MgCl_2$. F) Gap-filling reactions were performed as in C) but using a 2 nt gapped substrate containing the oligonucleotides P15, T17 and DG2-P. When indicated, the dNTP complementary to the first (dCTP), second (dGTP) or both templating bases were added (10 nM). G) NHEJ reactions were performed as in E) but using one set of substrates that gives shape to a 2 nt gap once joined: the labeled substrate contained the olignucleotides D3-C and D1, and the cold substrate, D4-AC and D2. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, each of the four dNTPs (10 μM) were added in the presence of 2.5 mM $MgCl_2$.

Their similar efficiency mimicks their behaviour in gap-filling. However, on incompatible ends that cannot be accurately joined (Fig. 52C, top panels), R387K was more efficient than wild type Pol*μ* in NHEJ, according to the enhanced terminal transferase and reduced activation barrier of this mutant (Fig. 51). On the other hand, on incompatible ends that can be accurately joined by the wild-type Pol*μ* (Fig. 52C, bottom panels), i.e. the reaction is preferentially template-directed (ddC insertion), mutant R387K showed an error-prone behaviour, as perhaps expected from its excessive terminal transferase activity. These results suggest that $Arg^{387}$ residue, which acts as a brake for terminal transferase activity in Pol*μ*, could avoid generation of unnecessary variability during NHEJ. Mutant R387A, on the other hand, is completely negative not only in terminal transferase activity (Fig. 47B), but also in any of the NHEJ substrates tested (Fig. 52B&C). This set of results indicate that not only the "brake" function of this residue (when interacting with the primer strand) is important, but also its secondary interaction with the -3 position of the template strand (that can still be happening in the R387K mutation) is strictly necessary for NHEJ catalysis in the case of substrates with a discontinuous template strand. Figure 52D (left panel) shows a representation of the initial step of a TdT reaction, in which the 1nt 3'-protruding DNA substrate is bound in a non-productive fashion, with its 3'-terminus in the incoming nucleotide position. Residues $His^{329}$ and $Arg^{387}$ are oriented as in TdT:ssDNA complex: $His^{329}$ does not stablish any interactions, while $Arg^{387}$ is contacting the primer strand in its "brake" position. In the central panel, the translocation of the primer has occurred, helped by arrival of the incoming nucleotide and by the new interactions stablished by $His^{329}$, with both the nucleotide and the primer-*terminus*; $Arg^{387}$ has finally allowed this movement by releasing the primer strand and is now contacting the -3 position of the template strand. In the last panel of figure 52D the main difference is the presence of a second DNA end, as it would happen in a NHEJ reaction. The binding of this template/downstream molecule would help the translocation of the primer strand and the release of the "braking" interaction mediated by $Arg^{387}$, through the stablishment of new interactions with the protein, that will be discussed later.

## 2. The art of binding NHEJ substrates.

### 2.1 NHEJ-specific Pol*μ* residues acting as ligands of the priming end: $Lys^{249}$, $Arg^{253}$, $Arg^{416}$.

During NHEJ reactions of 3'-protruding ends, Pol*μ* has to be able to interact with three substrates at once, i.e. the incoming nucleotide and the two DNA ends that need to be joined. In order to maintain all the substrates in proper register for polymerization to take place, Pol*μ* has to stablish a number of interactions with each of them. This part is devoted to the study of the contacts between the polymerase and the DNA substrate acting as primer. We have already discussed the importance of one residue involved in the interaction with the 3'-terminus of the primer, $His^{329}$, but we found three other residues in Pol*μ* that make contacts with the primer strand: $Lys^{249}$, $Arg^{253}$ and $Arg^{416}$, as shown in figure 53A. Figure 53B shows the level of conservation of each of these three residues among the members of the X family: the positive charge of $Lys^{249}$, that contacts the -5 position of the primer, is always conserved, except in the case of Polß in which a serine occupies this position; in the case of $Arg^{253}$, which interacts with the -4 position of the primer, the contact is maintained in all members except for Polλ, that has a methyonine; $Arg^{416}$, that interacts with the 3'-terminus of the primer, is completely invariant in all the members of the family. We mutated each of them to alanine in order to abolish their function. None of the individual mutations affected gap-filling, with the exception of R416A, that showed a slight reduction in activity (Fig. 53C). On the other hand, when these mutants were tested for terminal transferase activity, all of them showed a strong reduction in the level of untemplated incorporations (Fig. 53D). This phenotype was expected since these contacts with the primer strand are fundamental in the case of a substrate lacking a complete template strand and a downstream strand that could provide with a stabilizing 5'-P group. As expected, when we tested these mutants for NHEJ activity, all of the mutations completely abolished activity on non-complementary ends (Fig. 53E, bottom panels), but both K249A and R253A maintained a certain level of activity on complementary substrates (Fig. 53E, top panels). Given the importance of these residues for stabilizing the primer strand, we decided to test the

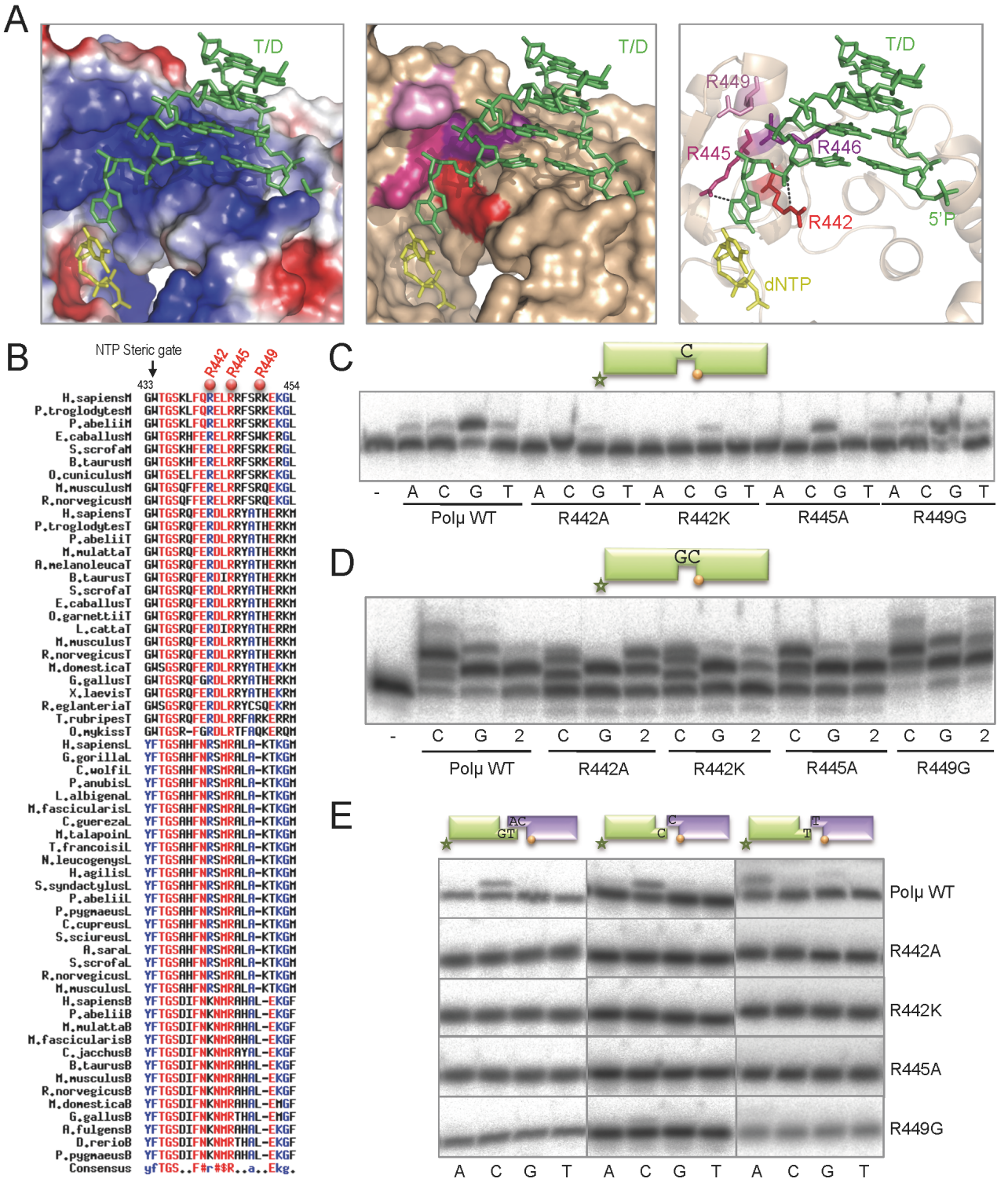



**Figure 54. The "arginine-helix": residues implicated in the binding of the template strand.**
A) Different respresentations of the Pol*μ* ternary complex structure showing the region of helix N: electrostatic surface (left panel) showing the high amount of positive charge, color-coded surface showing the positions of the four arginines present in helix N ()middle panel), and cartoon representation with the four arginines shown in sticks (right panel) and colored as in the previous panel. Incoming dNTP is shown in yellow and DNA substrate is shown in green. B) Multiple sequence alignment showing the zone corresponding to the helix N in Pol*μ* (M), TdT (T), Polλ (L) and Polß (B) from different species. Numbers indicate the residues in the human Pol*μ* sequence. The three arginines selected for muatagenesis are indicated with red dots. C) Gap-filling reactions were peformed as described in Materials and Methods with the indicated proteins (25 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. When indicated, dNTPs were added separately at 10 nM in the presence of 2.5 mM $MgCl_2$. After electrophoresis, labeled fragments were detected by autoradiography. D) Gap-filling reactions were performed as in C) but using a 2 nt gapped substrate containing the oligonucleotides P15, T17 and DG2-P. When indicated, the dNTP complementary to the first (dCTP), second (dGTP) or both templating bases were added (10 nM).. E) NHEJ reactions were performed with 200 nM of the indicated proteins and using four sets of substrates: the labeled substrates were formed by hybridization of 1G with 1D-NHEJ or D3-C with D1, and the cold substrates, by hybridization of either 2G with 2D-NHEJ or D4-C with D2. The orange balls indicate the presence of a 5'-P group in the downstream strand of the substrate. When indicated, each of the four ddNTPs (10 *μ*M) were added in the presence of 2.5 mM $MgCl_2$.

mutants for dislocation of the template during polymerization on a 2-nt gapped-substrate. In this case, mutants maintained the "dislocator" phenotype, incorporating preferentially the nucleotide complementary to the second templating base, even in the presence of both nucleotides (Fig. 53F), but no longer showed a similar level of gap-filling activity to that of the wild type Po*μ*. On the contrary, mutants were highly affected, specially R253A and R416A. This evidence is in accordance with the "dNTP selection"-mediated dislocation model, in which the polymerase has to produce a 1 nt gap by flipping-out the first templating nucleotide, in a kind of "scrunching" of the template strand (See scheme 1 in figure 53E). This same result can also be achieved by flipping out the -2 nucleotide, since the primer can be realigned as shown in scheme 2: a T:G mismatch, almost a Watson & Crick base-pair, would be formed. The lack of any of these positively charged residues implicates that Pol*μ* is not able to pull the primer strand and, therefore, is incapable of locating it close to the 5'-P of the downstream strand, the other important constraint that this polymerase encounters, either when filling-in a gap, or, more importantly, during the first step of NHEJ. To test this last hypothesis, we assayed the activity of the mutants on a complementary NHEJ substrates that once joined lead to the formation of a 2-nt gap that needs to be filled-in (Fig. 53G, see scheme). Wild-type Pol*μ* was able to use both bases

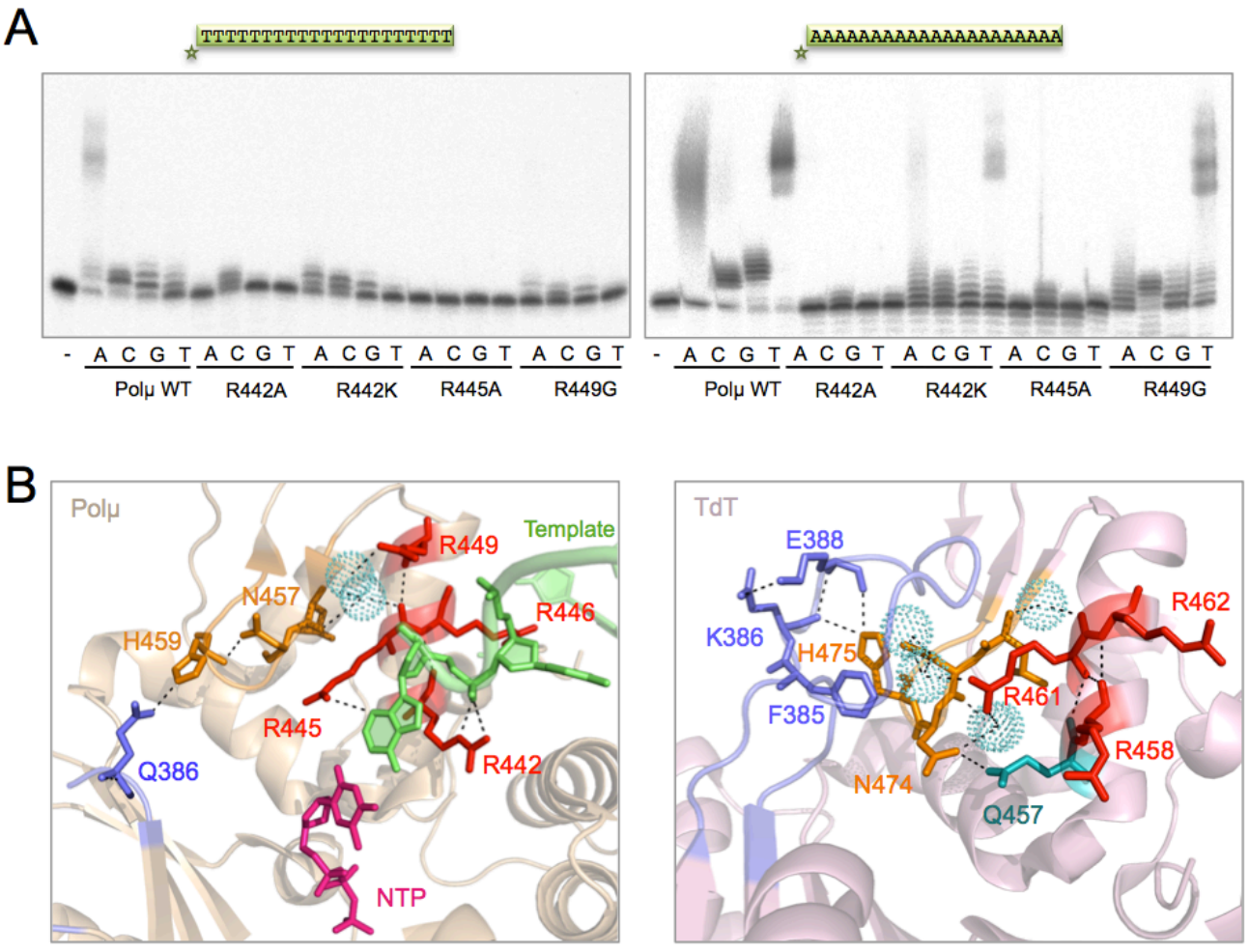



**Figure 55. The "arginine-helix": involvement in the terminal transferase activity of Pol*μ* through correct positioning of Loop 1.** A) Terminal transferase activity assay with the indicated proteins (600 nM) using two different homopolymeric substrates (polydT and polydA) and each of the four dNTPs (100 *μ*M). Reactions were performed for 30 min at 37ºC. After electrophoresis, labeled fragments were detected by autoradiography. B) Cartoon representations of the ternary complex of Pol*μ* (2IHM, left panel) and the TdT apoenzyme (1JMS, right panel). Loop 1 is shown in blue cartoon with selected residues involved in interactions shown in sticks, the thumb mini-loop is shown in orange with selected residues shown in sticks, arginines from the helix N are shown in red sticks, water molecules and othr residues involved in the network of interactins are shown in light teal. In the case of Pol*μ*, the incoming nucleotide is shown in dark pink and the template strand is shown in green.

as templates, inserting dT in front of the first templating base (A) and dG in front of the second (C). In each case, a different mode of template strand scrunching needs to occur, as shown in figure 36G, schemes 1 and 2. As expected, mutant R416A was unable to polymerize on this kind of substrates. Strikingly, mutants K249A and R253A only inserted dT when confronted with this substrate, meaning that the presence of this residues is what renders Pol*μ* able to promote the scrunching of the templating A (Fig. 53G, scheme 1), and thus, to dislocate the template strand in a NHEJ context. All this evidence shows the importance of these residues during NHEJ, firstly due to the non-continuous nature of the substrate that invokes the need for a tighter binding of the primer strand, and secondly to the highly probable formation of gaps longer than 1 nt and also distortions in the connection, such as mismatches or flipped-out nucleotides, that need to be accepted by the polymerase.

### 2.2 Residues implicated in the binding of the template strand: Arg$^{442}$, Arg$^{445}$, Arg$^{449}$.

After analyzing the NHEJ-specific Pol*μ* residues acting as ligands of the primer strand, we next focused on a positively charged $\alpha$-helix in human Pol*μ* which contains four arginines (Arg$^{442}$, Arg$^{445}$, Arg$^{446}$ and Arg$^{449}$) oriented towards the negatively charged phosphate backbone of the template strand (Fig. 54A). Arg$^{442}$ and Arg$^{445}$ (positive charges present in Pol*μ*, TdT, Polλ and Polß enzymes from different species) are analogous to Arg$^{514}$ and Arg$^{517}$ in human Polλ (Fig. 54B), which trigger both the DNA motion and the thumb loop motion (Foley et al., 2006), and Arg$^{517}$ also seems to control fidelity (Foley and Schlick, 2009), at least by *in silico* simulations. In human Polß, Arg$^{283}$, analogous to Polλ Arg$^{517}$, is also important for fidelity (Beard et al., 1996; Osheroff et al., 1999; Werneburg et al., 1996). Pol*μ* Arg$^{446}$ is not conserved in Polß or Polλ, where an alanine occupies the corresponding location, but it is conserved in TdT (Fig. 54B). Arg$^{449}$, on the other hand, is only conserved among Pol*μ*s of different species, and not in the other three members of the family. Arg$^{442}$ and Arg$^{445}$ in Pol*μ* interact with the DNA template through a series of hydrogen bonds and stacking interactions (Fig. 54A, right panel), in a similar manner to the interactions stablished in Polλ. This may suggest the importance of Arg$^{442}$ and Arg$^{445}$ in maintaining the active (ternary) form of Pol*μ*, *via* stabilization of the DNA template, specially in those situations in which the template strand is discontinuous. Arg$^{445}$ may participate in the active-site assembly when an incorrect nucleotide exists at the active site, as suggested by the similar fidelity checking function of Arg$^{517}$ in Polλ or Arg$^{283}$ in Polß. Arg$^{442}$ and Arg$^{445}$ are also conserved in TdT, which only catalyzes on single-stranded DNA with no template strand, and therefore these residues may also be important for interactions with other portions of the DNA or protein motifs in the polymerase. In a recent publication (Li and Schlick, 2010), analysis of molecular-dynamics simulations of Pol*μ* in different conditions showed that Arg$^{445}$ affects the conformation of Loop 1, and thus may be important for maintaining the Loop 1-DNA interactions that are crucial for template-independent synthesis (Juarez et al., 2006). Taking all these observations into account, we decided to prepare the mutants R442A, R442K, R445A and R449G in order to determine how these residues contribute to the function of Pol*μ*.

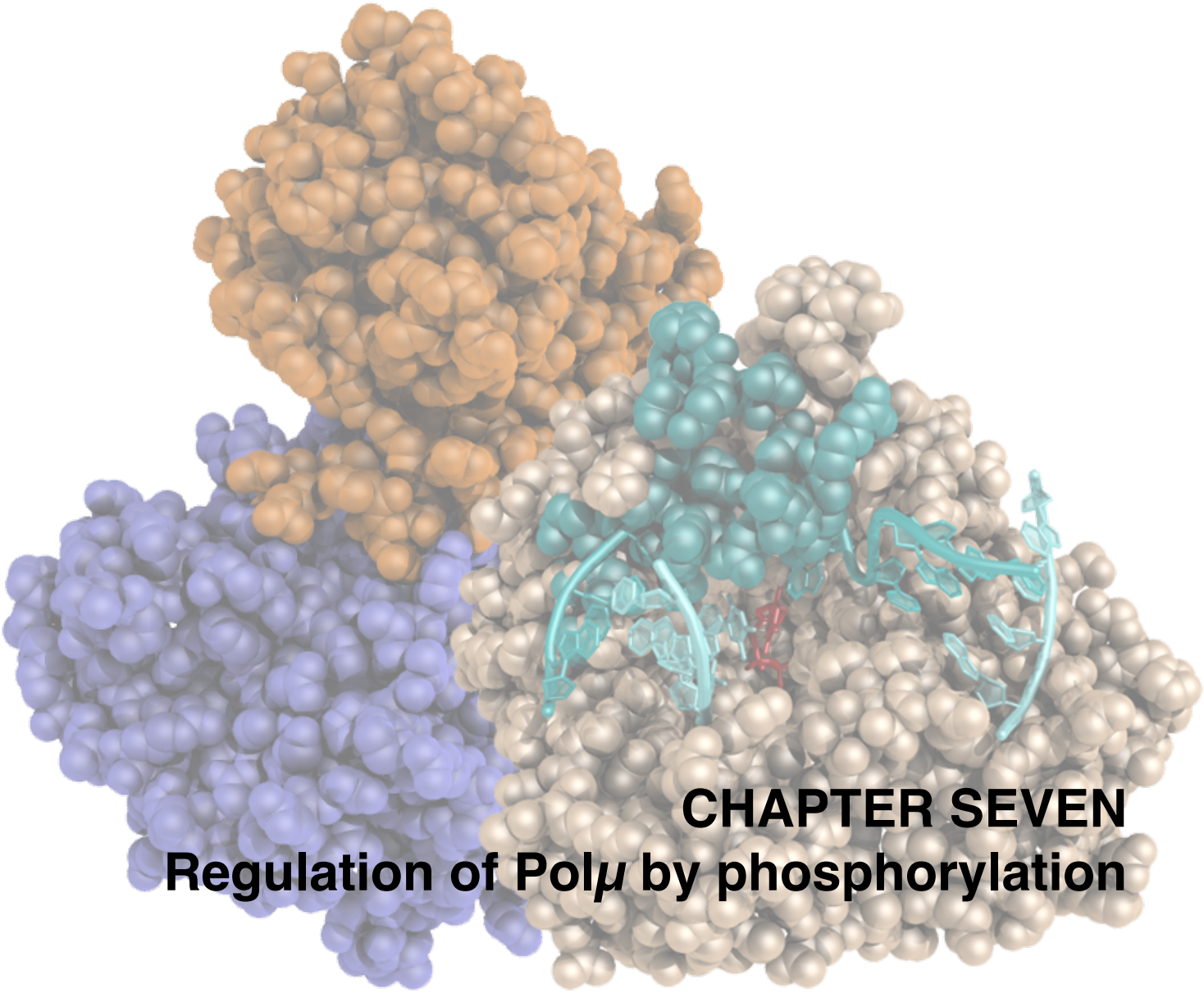
Figure 54C shows the activity of wild type Pol*μ* and the mutants during polymerization on a 1 nt gap, and, as expected, mutants in residues Arg$^{442}$ and Arg$^{445}$ were affected in their activity levels, but not in fidelity. Mutant R449G displayed wild type activity on this substrate. Polymerization on 2 nt gapped substrates was also lower except in the case of R449G (Fig. 54D), but the striking result is that mutations in residue Arg$^{442}$ seem to affect the ability of Pol*μ* to dislocate the template strand, since they showed a lower level of incorporation of the nucleotide complementary to the second base of the gap, and the mutants also showed a higher level of the +2 extension product when given both nucleotides at once (Fig. 54D). On this substrate, mutant R449G showed a very strong increase in the activity producing +3 and +4 extension products (Fig. 54D).

We then tested substrates with discontinuous template strands to check our hypothesis that the template-stabilizing function of these residues could be decisive when confronting NHEJ substrates. As shown in figure 54E, all the mutants were completely negative on these reactions, both in the case of compatible and non-compatible ends, the latter containing either a strong templating base (C) or a

weak template (T), situation in which the terminal transferase activity of Pol*μ* can contribute to create the connection between the two ends. These results clearly emphasize the importance of not only a strong binding to the DNA substrates, that still could be achieved in these mutants throughout binding to the 5'-P group in the downstream strand, but also the need for a perfectly orchestrated synapsis, in which all the elements (two DNA ends and incoming nucleotide) must be in proper register for catalysis.

In order to analyze if these residues are only implicated in stabilization of the template strand or whether they could be implicated in interactions with other DNA substrates or amino acid motifs in the polymerase (mainly the motifs implicated in the terminal transferase activity that operates on non-compatible ends with weak templating bases) we tested our mutants for polymerization on homopolymeric single-stranded DNA substrates (poly-dT and poly-dA). Strikingly, mutants R442A and R445A showed very low or undetectable levels of terminal transferase activity on both substrates, in the presence of any of the four dNTPs, in comparison to the wild-type Pol*μ* (Fig. 55A). On the other hand, mutant R442K, in which the charge of the residue is conserved but not its shape or length, and R449G, which is the residue located further away from the substrate, still displayed some terminal transferase activity.

A plausible explanation for the strong phenotype of mutants R442A and R445A in a reaction not involving a template strand, could be that these residues might be implicated in stabilizing Loop 1 as a template-mimicking structure, as already predicted in the molecular dynamics studies for $Arg^{445}$ (Li and Schlick, 2010). In agreement with this hypothesis is the observation that $Arg^{442}$ and $Arg^{445}$ are conserved in TdT ($Arg^{458}$ and $Arg^{461}$), and even though in the crystal structures available there is no direct interaction between any of these arginines and Loop 1, they are forming part of a network of interactions probably affecting the final position and intimate orientation of Loop 1: as shown in the previous chapter, there is a second loop located in the thumb subdomain of TdT and Pol*μ* which is implicated in stabilizing Loop 1 when this motif needs to be in the position of the missing template strand. Interestingly, Pol*μ* residues $Arg^{446}$ and $Arg^{449}$ interact through water molecules with residue $Asn^{457}$ from the "thumb" loop, which in turn interacts with $His^{459}$ that is contacting $Glu^{386}$, at the basis of Loop 1 in Pol*μ* (Fig. 55B, left panel). It is worth noting that in TdT a similar network is also found, as shown in figure 55B, right panel.



# CHAPTER SEVEN
# Regulation of Pol*μ* by phosphorylation

Regulation of the members of the family X starts at the transcriptional level, existing a differential tissue expression. This is specially true for TdT, which is only expressed in primary lymphoid organs, mainly in thymus and bone marrow, from the birth on but not in the embryo, stage at which Pol*μ* performs TdT functions (Gozalbo-Lopez et al., 2009). Pol*μ* is also regulated at this basic level, being preferentially expressed in peripheral lymphoid tissues (lymph nodes and spleen), in agreement with the fact that a large proportion of the human ESTs for Pol*μ* correspond to germinal center B cells (Dominguez et al., 2000). Polλ is highly expressed in testis, suggesting a role of this polymerase during the DNA repair processes associated with meiosis. Polß, on the other hand, is widely expressed, as expected from its function as a housekeeping DNA repair gene.

In the case of regulation through post-translational modifications, the ones affecting Polß include acetylation by p300 at an amino acid critical for the 5′dRP lyase function: $Lys^{72}$. Acetylation at this site blocks formation of the Schiff-base intermediate and results in abrogation of the 5′dRP lyase activity of Polß but leaves its gap-filling and DNA binding functions unaffected (Hasan et al., 2002). Evidence for a phosphorylated form of Polß has been reported, and this modified form exhibited higher BER activity than the unphosphorylated Polß (Kotake et al., 2002; Luo et al., 2007). The site of phosphorylation is not known and it is also unclear whether phosphorylation impacts catalytic activity or protein–protein interactions. Recently, Polß was found to form a complex with and is modified by protein arginine methyltransferase 6 (PRMT6), resulting in methylation of $Arg^{83}$ and $Arg^{152}$. Methylation of these residues stimulated processive polymerase activity and thus a preferential shuttle of repair through long-patch BER (El-Andaloussi et al., 2006). More recently, Polß was reported to be methylated (*in vitro* and *in vivo*) by PRMT1 but methylation was at amino acid residue $Arg^{137}$ (El-Andaloussi et al., 2007), not affecting DNA polymerase activity or 5'dRP lyase activity but diminishing the interaction with PCNA.

Concerning regulation of protein stability, TdT has been recently shown to be a target for proteasome degradation after interaction with BPOZ-2 (Maezawa et al., 2008). BPOZ-2 functions as a substrate-specific adaptor for CUL3, which in turn is one of the seven members of the Cullin-Rbx-type of E3 ubiquitin ligases. E3 ligases form part of the ubiquitin-proteasome machinery that directs target proteins to 26s proteasome degradation. BPOZ-2 expression is activated by the phosphatase and tensin homologue (PTEN), a tumor suppressor. Since BPOZ-2 over-expression inhibits cell cycle progression during the G1/S transition, BPOZ-2 may be involved in the growth suppressive effect of PTEN (Unoki and Nakamura, 2001). TdT, BPOZ-2, and CUL3 have been shown to form a ternary complex *in vivo*, after which TdT is ubiquitinated only within the nucleus and degraded by the 26S proteasome (Maezawa et al., 2008). Regulation of protein stability also applies to Polß: as recently described, stability of BER enzymes, including Polß, increases after formation of a repair complex on damaged DNA; proteins not involved in this repair complex are ubiquitylated by the E3 ubiquitin ligase CHIP and subsequently rapidly degraded (Parsons et al., 2008). More recent results from Dianov's laboratory indicated that DNA damage-associated monoubiquitylation of Polß precedes polyubiquitylation by CHIP, and identified Mule (ARF-BP1/HectH9) as the E3 ubiquitin ligase that is able to monoubiquitylate Polß (Parsons et al., 2009) and whose activity is regulated by ARF in response to DNA damage (Chen et al., 2005).

It has been shown that stability of Polλ protein is regulated in a cell cycle dependent manner (Wimmer et al., 2008). Cell cycle progression is regulated by a family of cyclin-dependent kinases (Cdks) which phosphorylate and activate proteins that execute events critical to cell cycle progression. For activity, Cdks require association with a cyclin and phosphorylation by a Cdk activating kinase (CAK) at a conserved threonine residue (Morgan, 1997). Each phase of the cell cycle is characterized by the expression of different Cdk/cyclin complexes that phosphorylate and regulate downstream substrates. In vertebrates, Cdk4/6/cyclin D complexes are active throughout the G1 phase, Cdk2/cyclin E at the G1/S boundary, Cdk2/cyclin A during S phase and Cdk1/cyclin A and Cdk1/cyclin B during the G2/M transition. Studies using knockout mice (Aleem et al., 2004; Geng et al., 2003; Ortega et al., 2003) revealed that neither Cdk2 nor cyclin E is essential for mitotic cell division. Cyclin E seems to be important for the control of endoreplication and for quiescent cells which re-enter in cell cycle.

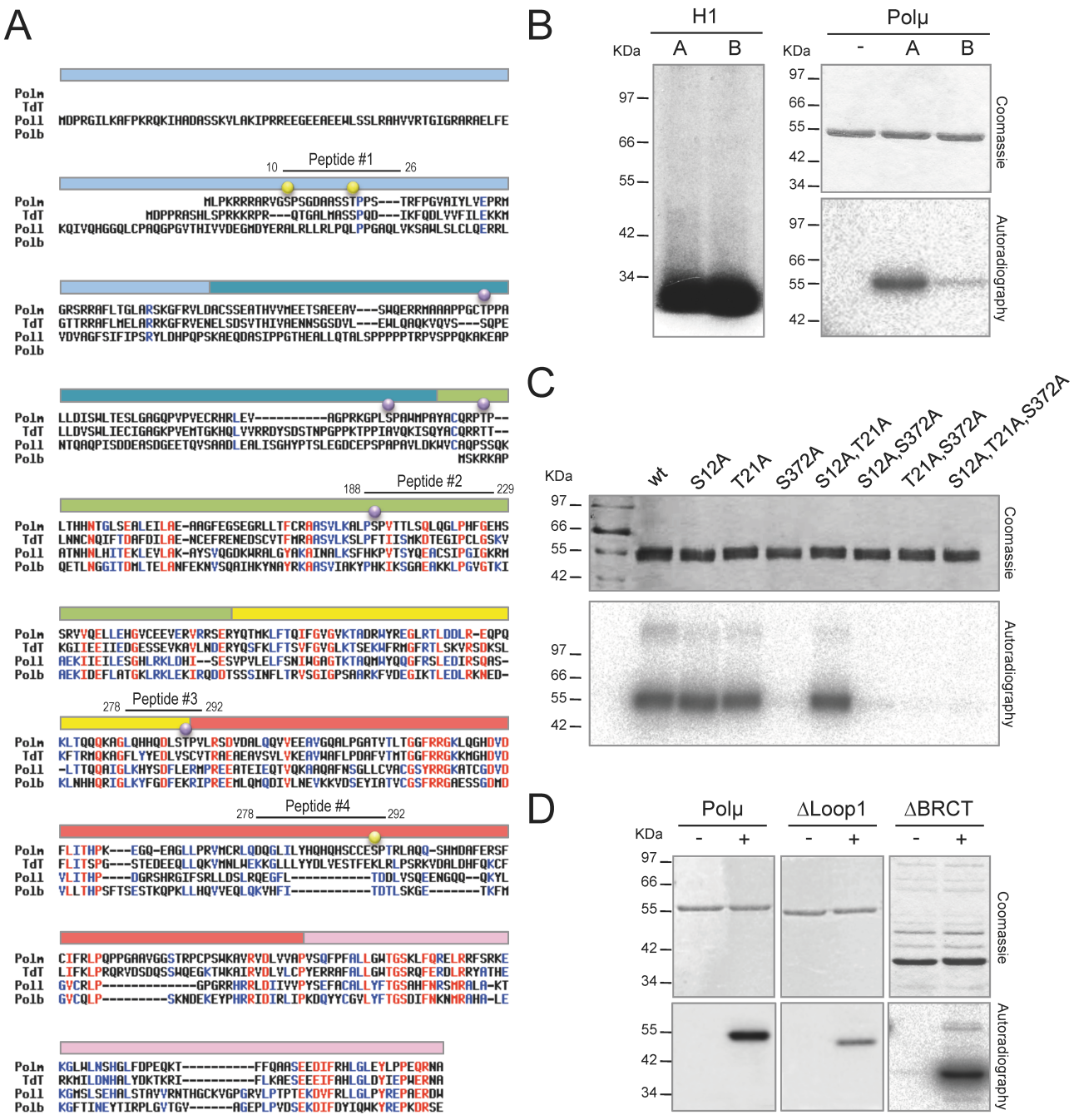



**Figure 56. Regulation of Pol*μ* through phosphorylation: residues modified *in vitro*.**
A) Sequence alignment of the four X family members in humans, Pol*μ* (M), TdT (T), Polλ (L) and Polß (B), showing the two regions in which the mutations were made. The peptides detected in the mass espectrometry analyis are indicated with black bars. Residues included in Cdk consensus sequences are indicated with purple and yellow dots. Residues with a yellow dot were phosphorylated in vitro. The colored bars above the sequence indicate the different domains of the tertary structure: light blue for the BRCT domain, teal for the Ser/Pro domain, green for the 8 kDa domain, yellow for the fingers, red for the palm and pink for the thumb. Invariant residues are shown in red, conservative in blue. B) *In vitro* Cdk phosphorylation assay of histone H1 (100 ng) and Pol*μ* (600 ng) was carried out as described in Materials and Methods. Two different Cdk complexes were used in this assay, Cdk2/cyclin A (A) and Cdc2/cyclin B (B). The amount of the different proteins was visualized by Coomassie staining, the labeled portion was detected by autoradiography. The molecular weigh markers are indicated in the left. B) *In vitro* Cdk phosphorylation assay of the wild-type and mutant versions of Pol*μ* (600 ng), with the Cdk2/cyclin complex. The amount of the different proteins was visualized by Coomassie staining, the labeled portion was detected by autoradiography. The molecular weigh markers are indicated in the left. D) *In vitro* Cdk phosphorylation assay of the wild-type and deletion mutant versions of Pol*μ* (600 ng), performed with the Cdk2/cyclin complex. The amount of the different proteins was visualized by Coomassie staining, the labeled portion was detected by autoradiography. The molecular weigh markers are indicated in the left.

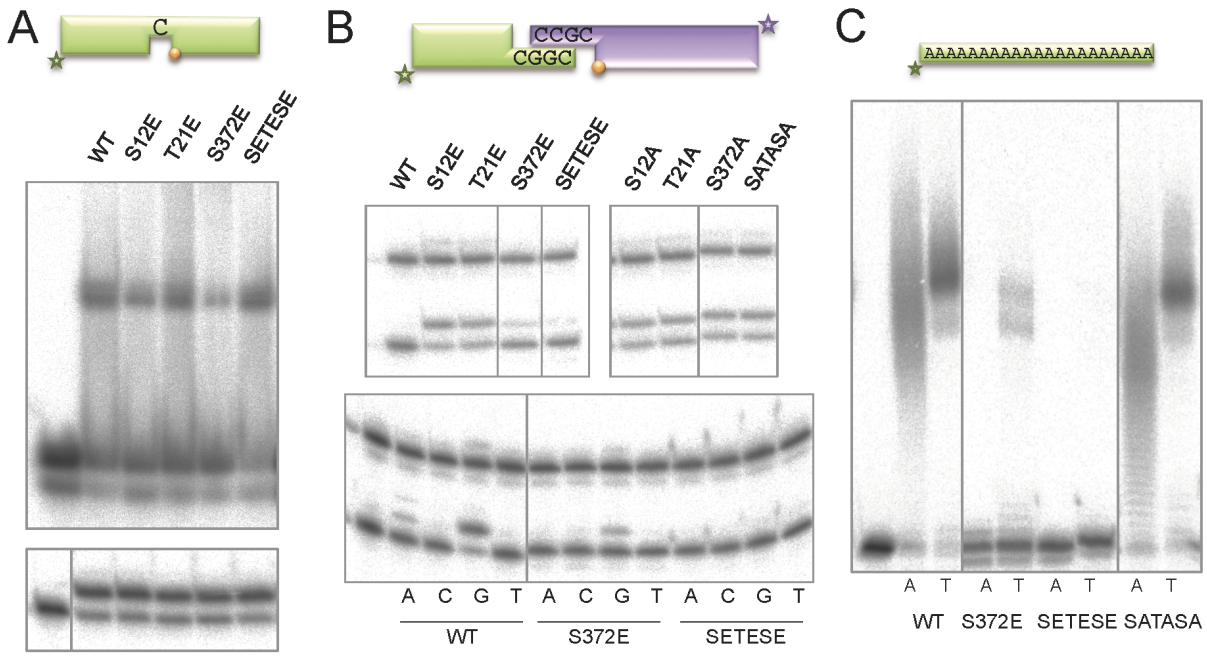



**Figure 57. Effect of the phospho-mimicking mutations: impaired terminal transferase and NHEJ.**
A) EMSA (top panel) was performed for the indicated proteins (200 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. After electrophoresis, gel was dried and the labeled fragments were detected by autoradiography. Gap-filling (lower panel) reactions were performed as described in Materials and Methods with the indicated proteins (25 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. When indicated, dNTPs were added separately at 10 nM in the presence of 2.5 mM $MgCl_2$. E) NHEJ reactions were performed with 200 nM of the indicated proteins and using one set of substrates: the short (green) substrate was formed by hybridization of D3-C with D1, and the cold substrate, by hybridization of D4-C with D2. The orange ball indicates the presence of a 5'-P group in the downstream strand of the substrate. In the top panel, only the correct dideoxynucleotide (ddG) was added (10 $\mu$M) in the presence of 2.5 mM $MgCl_2$. In the lower panel, each of the four ddNTPs (10 $\mu$M) were added in the presence of 2.5 mM $MgCl_2$. C) Terminal transferase activity assay with the indicated proteins (600 nM) using a homopolymeric substrate (polydT) and each of the four dNTPs (100 $\mu$M). Reactions were performed for 30 min at 37ºC. Labeled fragments were detected by autoradiography.

Furthermore, Cdk2 is essential in the regulation of the meiotic cell cycle, suggesting a novel tissue-specific function for cyclins and Cdks.

Previous work from Hübscher's laboratory showed that Polλ interacts with the cyclin-dependent kinase Cdk2 (Frouin et al., 2005). Moreover, Polλ was shown to be a substrate for phosphorylation by several Cdk/cyclin complexes *in vitro* and its proline-serine-rich domain is the target of phosphorylation by Cdk2/cyclin A. The phosphorylation of Polλ by Cdk2/cyclin A is decreased upon interaction of Polλ with PCNA. Their results also demonstrated that Polλ is phosphorylated *in vivo* and this phosphorylation is regulated during the cell cycle. A fourth phosphorylation site in Polλ, $Thr^{553}$, was described later (Wimmer et al., 2008), showing its importance for maintaining Polλ stability, as this DNA polymerase is targeted to the proteasomal degradation pathway through ubiquitination unless $Thr^{553}$ is phosphorylated. In particular, Polλ is stabilized during cell cycle progression in the late S and G2 phases. This most likely allows Polλ to correctly conduct repair of damaged DNA during and after S phase.

In the case of Pol*μ*, no post-translational modifications have been reported to date, and thus the possible mechanisms regulating either its activity or protein stability still need to be elucidated.

## 1. Phosphorylation of Pol*μ* by the S phase Cdk complex.

Taking into account the amino acid sequence similarities and close evolutionary relationship between Polλ and Pol*μ*, we decided to search for potential Cdk phosphorylation sites in the amino acid sequence of Pol*μ* (Fig. 56A). We found one serine, $Ser^{372}$, that forms part of a strict consensus CDK site (SPTR) located in a specific structural element, named Loop 1, only present in Pol*μ* and TdT (Juarez et al., 2006). Up to seven minimal consensus sites were also found, being $Ser^{12}$, $Thr^{21}$, $Thr^{95}$, $Ser^{134}$, $Thr^{147}$, $Ser^{191}$ and $Thr^{288}$ the putative phosphorylated residues for each site. In order to verify whether Pol*μ*

could be a substrate for Cdk-dependent phosphorylation, kinase assays with two different Cdk/cyclin complexes were performed in the presence of recombinant Pol$\mu$ and [g-$^{32}$P]ATP. As shown in Figure 56B, a strong [$^{32}$P]-labeled band of the expected molecular weight for Pol$\mu$ appeared only after incubation with the Cdk2/cyclin A complex (related to the S phase of the cell cycle), while only a light band appeared with the Cdc2/cyclin B complex (G2/M phases). Histone H1 was used as a positive control of phosphorylation by the two complexes. Identification of the phosphorylated band as Pol$\mu$ was confirmed by SDS–PAGE (Fig. 56B). The products of the phosphorylation assays were analyzed by mass spectrometry, where these four peptides were detected from the total Pol$\mu$ sequence: 10-VG**S**PSGDAASS**T**PPSTR-26, 188-ALP**S**PVTTLSQLQGLPHFGEHSSR-229, 278-AGLQHHQDLS**T**PVLR-292 and 354-LQDQGLILYHQHQHSCCE**S**PTR-375. The results indicated that the two peptides corresponding to residues Ser$^{21}$-Thr$^{21}$ and Ser$^{372}$ were phosphorylated, the first one including a double phosphorylation. In summary, these results suggest that Pol$\mu$ is a target for phosphorylation by Cdk2/cyclin A complex *in vitro*.

**2. Mutational analysis of Pol$\mu$ phosphorylation sites.**

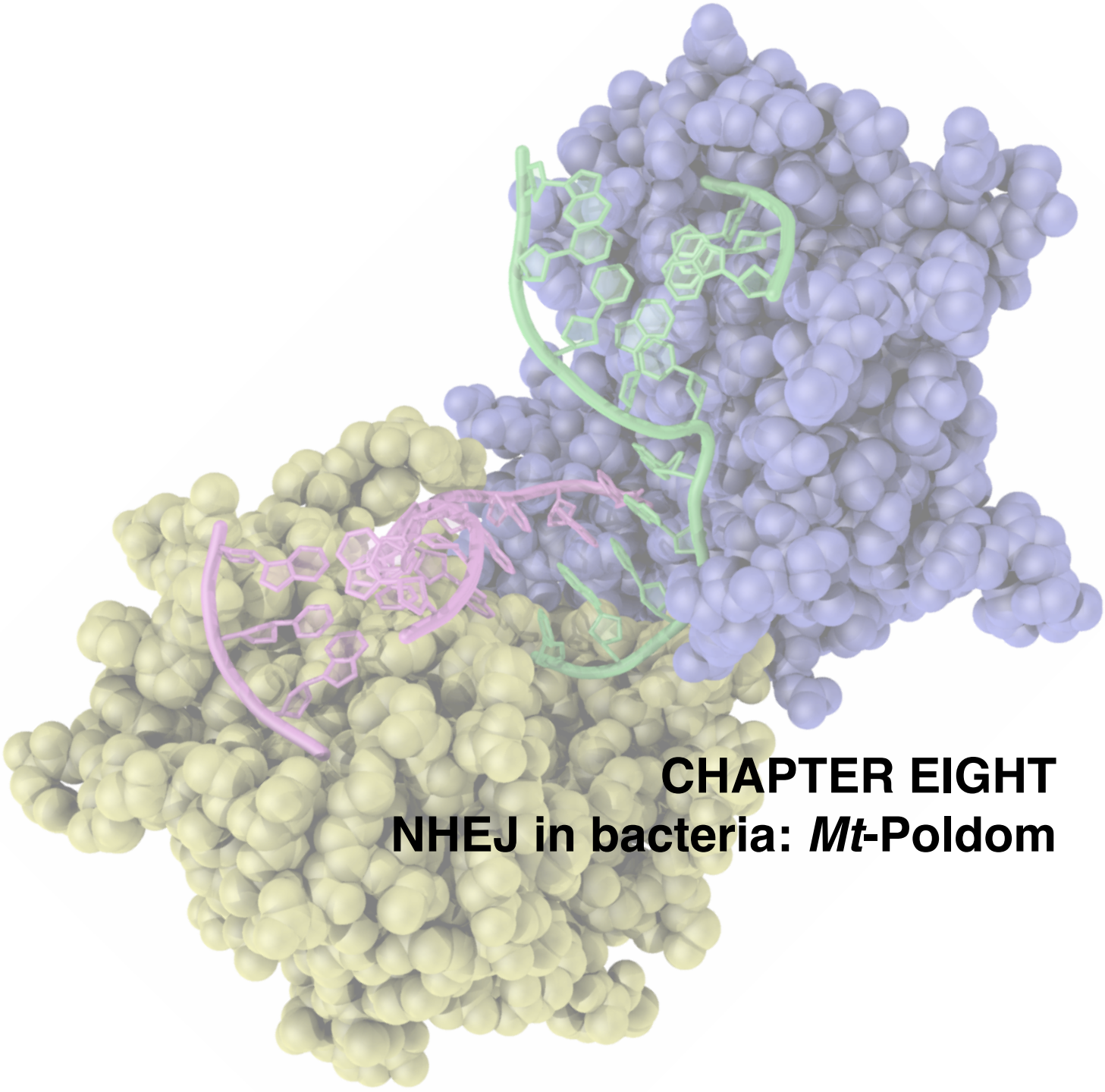
Once the capacity of Cdk2/cyclin A complex to phosphorylate Pol$\mu$ was ascertained, and three phosphorylation consensus sites had been validated, the three candidate residues were mutated to alanine individually and in combination (double, triple) by site-directed mutagenesis, and the mutant proteins were over-expressed in *E. coli* and purified. The level of phosphorylation of the different mutants was tested in an *in vitro* kinase assay with Cdk2/cyclin A and, as shown in figure 56C, the single mutants S12A and T21A displayed nearly the same phosphorylation level as the wild type Pol$\mu$; however, a decrease of more than 90% was achieved in the case of S372A (mutating the strict consensus site). Double and triple mutants showed that this residual phosphorylation was occurring at the other two sites located at the N-terminal portion of the protein. To confirm these results, we tested two other mutant versions of Pol$\mu$: Pol$\mu$-ΔBRCT, which would lack Ser$^{12}$ and Thr$^{21}$, and Pol$\mu$-Δloop1, which would lack Ser$^{372}$. As expected, when eliminating the strict consensus site located in the Loop 1, less than 30% of phosphorylation was detectable (Fig. 56D middle panel), while the levels were maintained when the other two sites were removed (Fig. 56D, right panel).

**3. Effect of phospho-mimicking mutations (glutamic acid) on Pol$\mu$ activity.**

In order to test the possible effect of the phosphorylation on the different activities performed by Pol$\mu$, and taking into account that the very low percentage of phosphorylation achieved in an *in vitro* assay did not allow comparison between the modified and non-modified forms of the enzyme, we mutated the target residues Ser$^{12}$, Thr$^{21}$ and Ser$^{372}$ to glutamic acid, a widely used approach to create a phospho-mimicking site due to the negative charge of the mutation introduced. Each single (S12E, T21E and S372E) and the triple mutant were over-expressed in *E. coli*, purified and tested in biochemical assays. We measured the binding (upper panel) and the activity (lower panel) of the mutants on gapped substrates and the levels reached were indistinguishable from that of the wild-type Pol$\mu$ (Fig. 57A). However, when tested for terminal transferase or NHEJ activity, all the mutants showed a decrease in the activity, being the triple mutation the most deleterious, with null activity on both assays (Fig. 57B&C). To check whether these alterations were due to the phospho-mimicking negative charge or to other structural modifications caused by the mutagenesis, we tested the triple alanine mutant in the same assays, and it behaved exactly like wild-type Pol$\mu$ both in terminal transferase and in NHEJ (Fig. 57B&C, last lane in each case). The single mutants were also analyzed to determine which of the phosphorylation events was responsible for this loss of specific activities, and although mutants S12E and T21A were partially affected, mutant S372E was the most affected in both activities (Fig. 57B&C), an anticipated result due to the location of Ser$^{372}$ in Loop 1, essential for the most distinctive activities of Pol$\mu$ as shown before (Juarez et al., 2006}; this PhD thesis). Phosphorylation of this residue could block the necessary movements of this region of the polymerase or hinder its interaction with the DNA substrate, either single stranded or double stranded with a discontinuous template. In the case of the

mutations in the N-terminal portion of Pol*μ*, a negative charge in this part of the BRCT domain could be interfering with the predicted interaction of this domain and the DNA substrate, already presented in chapter 1 of this PhD thesis.

In conclusion, this evidence suggests that Pol*μ* could be regulated *in vivo* at the functional level by phosphorylation events affecting mainly Loop 1. Therefore, its most distinctive activities would be turned off at specific cell-cycle phases (late S and G2), in which Pol*μ* promiscuity might be harmful to the cell.



# CHAPTER EIGHT
# NHEJ in bacteria: *Mt*-Poldom

In higher eukaryotes, the NHEJ pathway is critical for the repair of DSBs (Krejci et al., 2003). A functionally homologous repair system exists in many prokaryotes, where it is used to repair DSBs in stationary-phase and sporulating cells (Pitcher et al., 2007; Weller et al., 2002). The bacterial NHEJ complex is composed of two proteins, Ku and a multifunctional DNA ligase (LigD) (Bowater and Doherty, 2006; Della et al., 2004; Gong et al., 2005; Pitcher et al., 2007; Pitcher et al., 2006; Pitcher et al., 2005; Weller et al., 2002). In addition to a core ligase domain, LigD often possesses polymerase (PolDom) and nuclease domains (Bowater and Doherty, 2006; Della et al., 2004; Gong et al., 2005; Pitcher et al., 2007; Pitcher et al., 2006; Pitcher et al., 2005; Yakovleva and Shuman, 2006; Zhu et al., 2006). PolDom, a member of the archaeo-eukaryotic primase (AEP) superfamily (Della et al., 2004; Iyer et al., 2005; Lao-Sirieix et al., 2005), in turn has a variety of nucleotidyl transferase activities (Bowater and Doherty, 2006; Della et al., 2004; Gong et al., 2004; Pitcher et al., 2005) as well as the ability to generate template distortions and primer realignment (Pitcher et al., 2007; Yakovleva and Shuman, 2006). LigD from *Mycobacterium tuberculosis* (*Mt*-LigD) includes polymerase (*Mt*-PolDom) and nuclease domains that reside as N-terminal distal extensions of the ligase domain. *Mt*-PolDom possesses a remarkable variety of nucleotidyl-transferase activities including DNA-dependent RNA primase, terminal transferase and DNA-dependent DNA/RNA gap-filling polymerase activities (Della et al., 2004; Gong et al., 2004; Pitcher et al., 2005; Weller et al., 2002). *Mt*-PolDom (amino acid residues 1–300), when expressed in isolation, retained both DNA end-filling and primase activities (Pitcher et al., 2005). Previous work from the laboratory, in collaboration with Prof. Aidan Doherty, was focused on the study of several crystal structures of the polymerase domain from *M.*

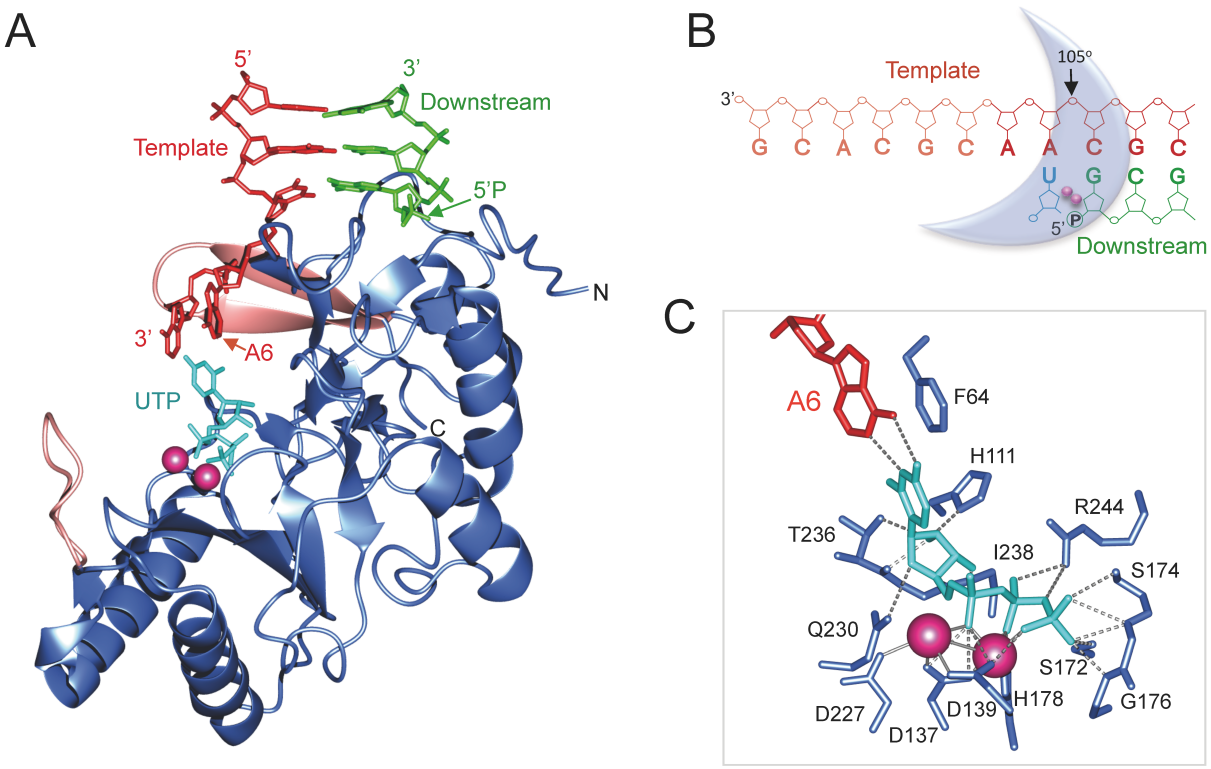



**Figure 58. Crystal Structure of the Preternary Complex of Mt-PolDom:DNA:UTP**
A) Ribbon representation of the preternary complex depicting the interaction of PolDom (blue) with the dsDNA and UTP. Loops 1 and 2 are highlighted in salmon pink showing their spatial relationship to the bound DNA and UTP molecules. DNA is colored green for downstream (D) strand and red for templating (T) strand. The bound UTP molecule is colored cyan and the cata- lytic manganese ions are pink. B) Schematic representation of Mt-PolDom binding to DNA and UTP. The unstructured 3'-overhanging DNA is shown as a faded color as it is unstructured. DNA chains follow the same color scheme as in (A). PolDom is represented by a blue crescent. The arrow indicates a kink angle (105º) between A6 and C5 of the template strand. C) A close-up view of the active site of Mt-PolDom highlighting the important stacking of A6 against Phe64 and the specific interactions of PolDom with UTP and UTP with DNA (A6). The DNA and protein monomer are colored as in A). Hydrogen bonds are shown in gray.

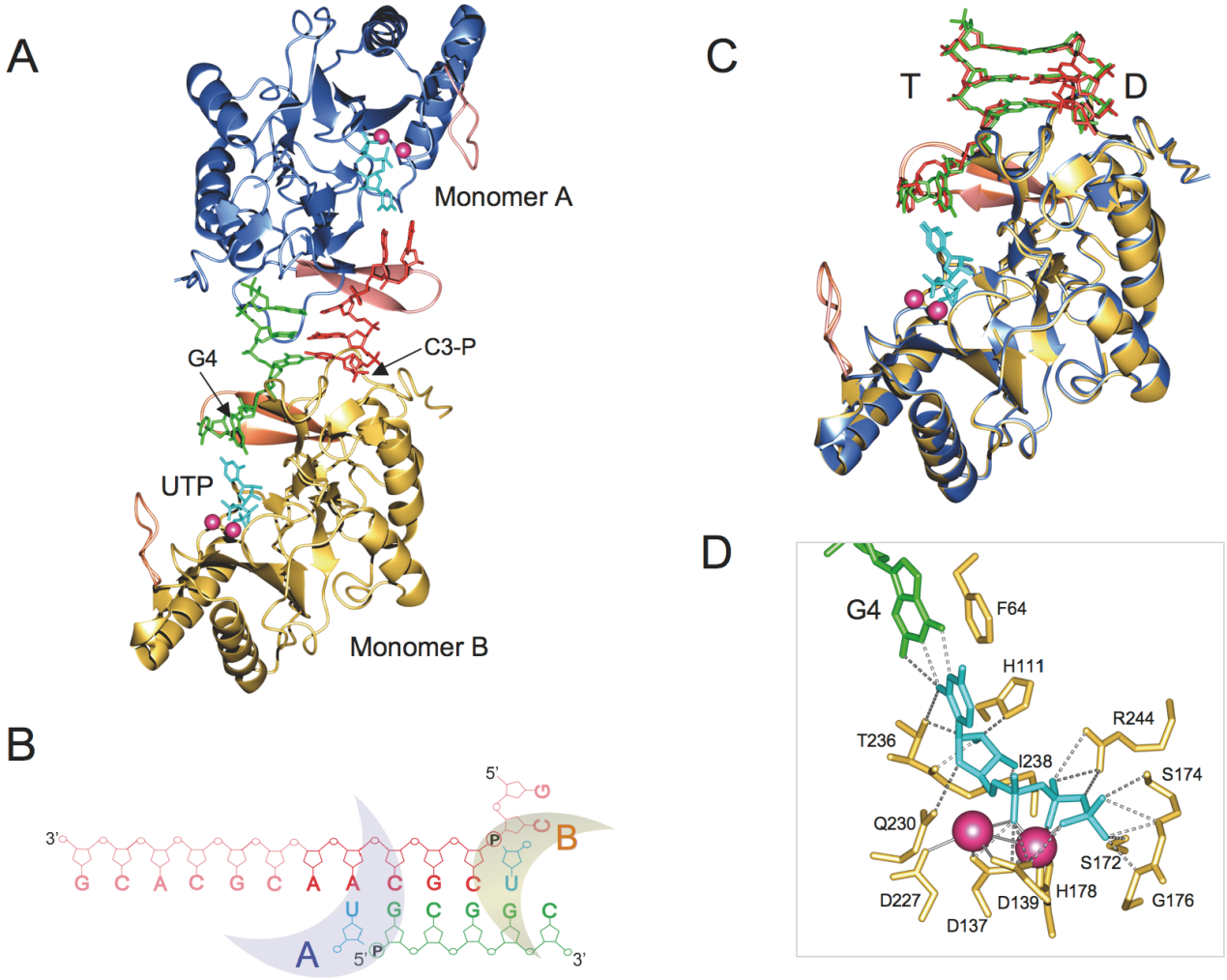



**Figure 59. Orientation of the second pre-ternary complex in the crystallographic complex.**
A) The second *Mt*-PolDom monomer B (gold) is shown in relation to the previously described monomer A (blue). Loops 1 and 2 are depicted in pink. The DNA acting as templating strand for monomer B is depicted in green, and the downstream strand in red. Complexed UTP and manganese ions are coloured cyan and deep pink, respectively. B) A schematic arrangement of protein monomers binding DNA and UTP as used in the crystallisation experiment. C) Superposition of the two pre-ternary complexes (monomers A and B). Colours are as before.D) A close-up view of the UTP bound in the second pre-ternary complex. Residues from monomer B, the UTP and manganese ions are coloured as before.

*tuberculosis*: the apo and the co-crystal structures with either GTP or dGTP bound in the active site (Pitcher et al., 2007), as well as the crystal structure of a NHEJ polymerase–mediated synaptic complex, which reveals a DNA-directed mechanism used by repair polymerases to induce synapsis of non-complementary ends through a dimeric arrangement (Brissett et al., 2007). Our collaborative effort went on to define and confirm biochemically a new crystal structure, as described below.

**1. Overall analysis of the pre-ternary complex crystal structure.**

The structural basis for the recognition of a 5'-phosphate on DNA by a mycobacterial NHEJ polymerase/primase has recently been elucidated (Brissett et al., 2007). Biochemical and structural studies have established that these NHEJ polymerases can coordinate end-synapsis, the connection between two DNA distinct molecules, suggesting that they may play a major role in remodeling of non-homologous termini prior to ligation. The structure of the LigD polymerase domain bound to DNA has established the molecular basis of DNA recognition and end-synapsis by the mycobacterial AEP-like NHEJ polymerases. *Mt*-PolDom was previously crystallized in complex with DNA consisting of a template strand (T; 13 bases) annealed to a downstream primer strand (D; 5 bases) containing a 5' phosphate, producing a dsDNA with a 3' overhang and a recessed 5'-P (Brissett et al., 2007). However, despite the inclusion of co-factors in the crystallization solution, the

synaptic complex lacked either bound nucleotide (ddTTP) or divalent metal ions. The absence of nucleotide in the active site may be explained by the low affinity of the active site for deoxyribonucleotides (Pitcher et al., 2005). Thus, this structure did not provide insights into the catalytic mechanism of these nucleotidyltransferases.

To obtain a ternary *Mt*-PolDom complex (*Mt*-PolDom-NTP-DNA), we soaked in UTP and manganese into equivalent crystals used previously to elucidate the structure of the synaptic complex (Brissett et al., 2007), resulting in a significant transformation of the space group from $P2_12_12_1$ to $P4_1$. The crystal structure was determined by molecular replacement using the structure of apo *Mt*-PolDom as the primary search model (Pitcher et al., 2005). The structure comprises amino acid residues 6-291, with no electron density observed for nine residues at the C-terminus. During refinement, two *Mt*-PolDom molecules were located in the unit cell bound to DNA. It was anticipated that the polymerase-mediated synaptic connection, present in the original structure, would be maintained in the UTP-bound *Mt*-PolDom-DNA complex. However, addition of cofactors induced major structural rearrangements within the unit cell, resulting in a change in the ratio of protein bound to DNA from 2:2 (in the synaptic complex) to 2:1 in the new complex, in which no synapsis is observed (Fig. 59A). The overall complex in the asymmetric unit consists of two distinct ternary-like complexes resulting from the binding of two polymerases (monomers A in blue and B in gold) to a single double stranded DNA molecule. Whereas the general fold of the two monomers is identical, and both contain a single base-paired UTP and two $Mn^{2+}$ ions bound in the active site (Fig. 59A), one *Mt*-PolDom monomer (A) is bound to the 3'-overhang terminus end, whilst the other monomer (B) binds to the opposing blunt end of the same DNA molecule. *Mt*-PolDom (monomer A) interacts with the recessed 5'-phosphate of the duplex strand (D). A similar interaction is made by monomer B, but *via* an internal phosphodiester (Fig. 59B). *Mt*-PolDom (A) orients the 3'-overhanging end of the templating strand (T), which is splayed out by ~105° (Fig. 58 A&B). Only two bases from the 3'-overhanging strand have observable density in the refined structure suggesting that the remaining six bases are highly mobile due to the loss of the stabilizing connection with the overhanging strand on the adjacent DNA molecule observed in the synaptic complex (Brissett et al., 2007). PolDom contains a single UTP bound in its active site, which forms a Watson-Crick base-pair with a templating adenine base (A6) with two metals ions also bound in close proximity (Fig. 58C). The 3'-OH group of the ribose moiety of UTP is hydrogen bonded to the amino nitrogen of I238. The 2'-OH group is hydrogen bonded to the side-chain hydroxyl and backbone carboxyl oxygen of T236, with a further hydrogen bond formed between the ∂N of H111 and the 2'-OH. Q230 interacts with the oxygen (O4') on the ribose moiety (Figs. 58C and 59C). The specific details of other nucleotide interactions, including contacts with the triphosphate tail, are described later.

**2. Formation of a second pre-ternary complex.**

The presence of a second PolDom:NTP:DNA complex (monomer B) was unexpected given that the blunt-ended duplex DNA lacks a terminal 5'-phosphate, a pre-requisite for DNA binding by these NHEJ polymerases (Brissett et al., 2007; Pitcher et al., 2005). Strikingly, two terminal base-pairs have been "peeled" open, allowing ingress of the polymerase into the blunt-ended region of the DNA, with three intact Watson-Crick base-pairs remaining (Fig. 59B). The positively charged phosphate-binding pocket (formed in part by $Lys^{16}$ and $Lys^{26}$) binds to an internal phosphate moiety (T-strand, base C3) mimicking the type of interaction with a terminal 5' phosphate observed in monomer A (Fig. 61). The O2 atom (of the phosphate group) hydrogen bonds with the primary amine of $Lys^{16}$ and the carbonyl oxygen of $Pro^{55}$. The O1 from the same phosphate group hydrogen bonds to the N group of $Asn^{13}$ and the primary amine group of $Lys^{26}$ (Fig. 61C&E). All these residues are invariant in the PolDom of bacterial LigDs (Fig. 63). This atypical phosphate recognition site enables monomer B to form a stable complex with the DNA that is almost identical in every way to the monomer A complex (Fig. 59).

**3. Templating base-pair interactions with UTP.**

In the current pre-ternary complexes, UTP is bound as a catalytically competent substrate in the active site of *Mt*-PolDom (Fig. 58). A major distinction of the

pre-ternary complexes from the previously described PolDom:NTP binary complexes (Pitcher et al., 2007) is the existence of a templating base, stabilized by a hydrophobic stacking interaction with $Phe^{64}$, which hydrogen bonds with the uracil base of UTP. In monomer A, UTP forms a Watson-Crick base pairing with A6 of the templating DNA strand (Fig. 58C). The base pair interaction occurs *via* N3 of UTP donating a hydrogen bond to N1 of the adenine base of A6 and O4 of UTP accepting a hydrogen bond from N6 of the adenine base of A6. In contrast to complex A, the UTP in monomer B forms a reverse wobble base-pairing arrangement with the templating guanine base (N1 of guanosine donates a hydrogen bond to O2 of UTP and O6 of guanosine accepts a hydrogen bond from N3 of UTP) on the 3' overhang of the downstream strand (Fig. 59D). Thus, the downstream primer strand from complex A acts as a templating strand for complex B. Likewise, the templating strand from complex A is behaving in part as the downstream strand for complex B.

**4. Additional protein-nucleotide contacts in the pre-ternary complex.**

In the pre-ternary complex, the factors that stabilize the triphosphate tail of UTP are very similar to those previously described (Pitcher et al., 2007). In addition to the interaction *via* the invariant $His^{178}$, the two manganese atoms, ligated by three conserved aspartate residues ($Asp^{137}$, $Asp^{139}$ and $Asp^{16227}$), contribute to maintaining the bound UTP nucleotide in a catalytically competent orientation (Figs. 58C and 59B). The γ-phosphate of the UTP forms dative bonds to the two $Mn^{2+}$ ions as opposed to hydrogen bonding with the side-chain of $Arg^{244}$, also observed in the PolDom:GTP crystal structure (PDB ID: 2IRX; Pitcher et al., 2007). In the PolDom:GTP binary complex, the triphosphate tail of the bound NTP is stabilized by a positively charged ridge, composed of $Lys^{16}$, $Arg^{52}$, $Lys^{175}$, $Arg^{244}$, $Arg^{246}$, which lines the edge of the active site pocket (Fig. 59B). In the pre-ternary complex, the side-chains of these residues (with the exception of $Arg^{244}$), whilst still maintaining the charged ridge, do not hydrogen bond with UTP. Moreover, the nucleotide in the pre-ternary complex interacts with the side-chains of $Ser^{172}$ and $Ser^{174}$ and makes backbone contacts with $Lys^{175}$ and $Gly^{176}$, acting as further ligands for the triphosphate tail of UTP (Figs. 58C and 59D). All these residues ($Asp^{137}$, $Asp^{139}$, $Ser^{172}$, $Lys^{175}$, $Gly^{176}$, $His^{178}$, $Asp^{227}$, $Arg^{244}$ and $Arg^{246}$), with the exception of $Lys^{52}$ and $Ser^{174}$, are invariant in the polymerization domain (PolDom) of bacterial LigDs (Fig. 63). The orientation of the nucleoside of the UTP in the pre-ternary complexes differs from nucleoside orientations in previously described PolDom:NTP binary complexes. For monomer A, the uracil base forms a Watson-Crick base pair with A6 of the templating strand. In monomer B, the uracil base forms a reverse wobble base-pairing with G4 of the downstream DNA strand (Fig. 59D). For these base-pairings to occur, the orientation of the ribose moiety of the nucleoside has to adopt a different orientation to that previously seen in PolDom:NTP binary complexes. Hydrogen bonding of the ribose is described in the main text. The most notable contact is that of $Gln^{230}$ with the oxygen (O4') on the ribose moiety of UTP, which occurs *via* a hydrogen bond with the N group of this residue. That specific glutamine residue is highly conserved or substituted by an arginine residue in LigDs (Fig. 63).

**5. Protein-DNA contacts in the pre-ternary complexes.**

As previously noted (Brissett et al., 2007), *Mt*-PolDom interacts with the DNA duplex predominantly *via* contacts with the recessed 5' phosphate moiety ($Asn^{13}$, $Lys^{16}$, $Lys^{26}$, $Arg^{53}$, $Pro^{55}$; the last four being invariant in LigDs; Fig. 63). The notable differences between the two DNA-bound PolDom structures (synaptic and pre-ternary) is the loss of the side-chain contact between $Ser^{71}$ and the DNA in the pre-ternary complex. $Gln^{104}$, which is absolutely invariant in LigDs (Fig. 63), maintains the hydrogen bond with the phosphate of C2 (downstream primer for monomer A, phosphate of G4 templating strand for monomer B) (Fig. 64). Further contacts with the template strand are depicted in figure 64A, and most of these are highly conserved among LigD members (see Fig. 463). The main-chain and side-chain atoms of $Glu^{65}$ (highly conserved/or substituted by a Gln in LigDs; see Fig. 63) are in non-bonded contact to A6 for monomer A and G4 for monomer B. Other contacts with the DNA, including $Lys^{66}$ and $Gln^{67}$ (to A6 for monomer A, to G4 for monomer B) are still maintained whereas the contact with $Thr^{88}$ is lost. As expected, the protein-DNA contacts made by

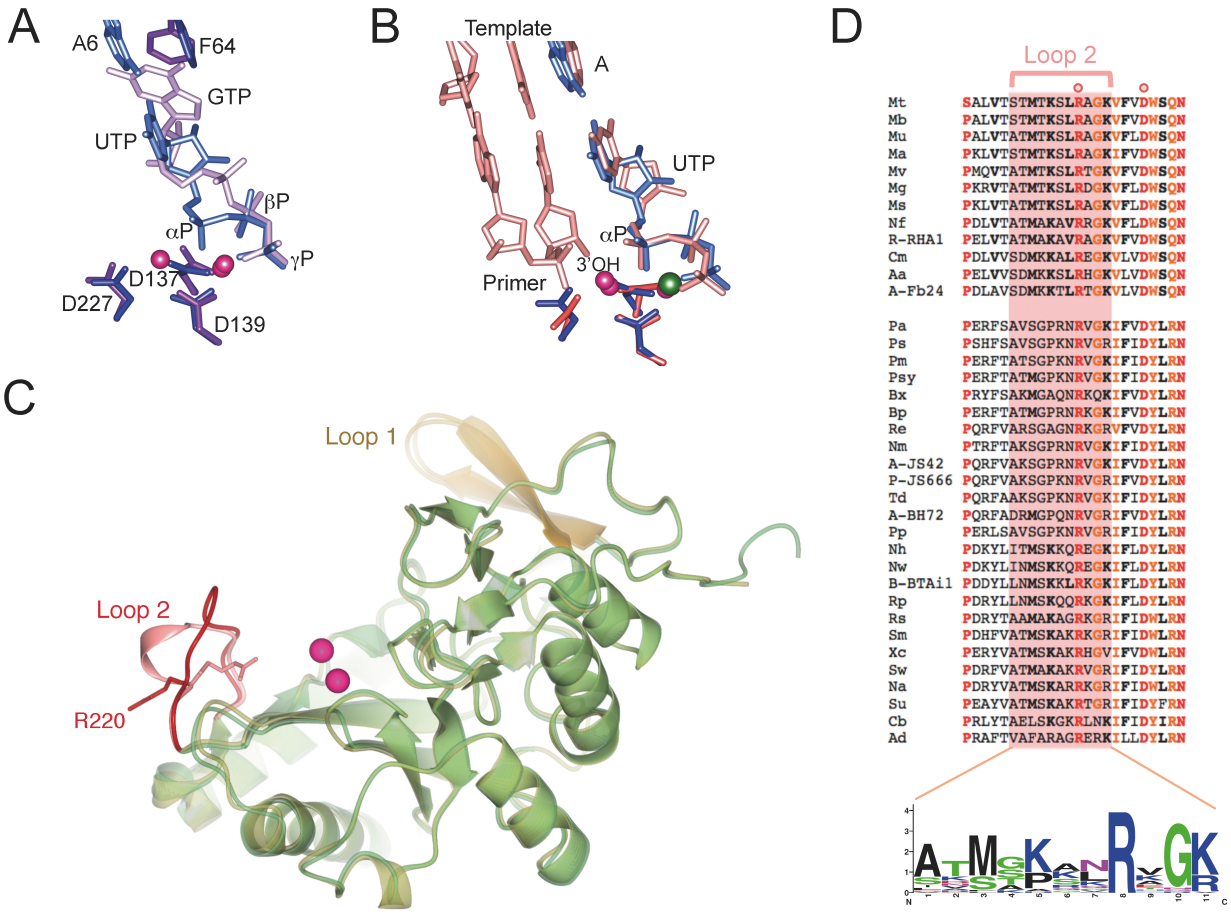



**Figure 60. Loop2: conservation and repositioning in the pre-ternary complex.**
A) Schematic view of the changes affecting nucleotide positioning, comparing the binary (Mt-PolDom:GTP; PDB ID:2IRX) complex (lilac) versus the preternary complex (blue). Metal ions (A and B) are indicated as pink spheres. Rotation of Phe$^{64}$ to establish alternative interactions either with the incoming nucleotide (GTP) or with the templating base (A6) is highlighted. B) UTP is in a catalytically competent conforma- tion in the preternary complex. Superposition of the catalytic site of Polλ (PDB ID: 2PFO) onto the catalytic site of the PolDom preternary complex. Catalytic aspartates from Mt-PolDom are colored dark blue with the UTP and templating base from the same structure colored light blue. Catalytic aspartates from the Polλ structure are colored red and the DNA, 20 -deoxyuridine 50 -alpha, beta-imido-triphos- phate (dUPnPP) colored salmon pink. $Mn^{2+}$ ions are colored pink, and $Mg^{2+}$ ions are colored green. (C) Ribbon representation highlighting the confor- mational changes associated with loop2, largely affecting the orientation of Lys$^{217}$ and Arg$^{220}$. Mt-PolDom structures corresponding to the binary (PDB ID: 2IRX) and preternary complexes were superposed. The two metal ions present in the preternary complex are indicated in deep pink. The region corresponding to loop2, and residue Arg$^{220}$, are indicated either in red (binary) or in pink (preternary). Loop1 is also depicted (yellow) to show the relative orientation of these structural elements. D) Amino acid sequence alignment of the PolDom region encompassing loop2 in the two main groups of bacterial LigD enzymes. Invariant or highly conserved amino acids are indicated in red and orange, respectively. Dots above the Mt-PolDom sequence indicate the catalytically relevant Arg$^{220}$ and the metal ligand carboxylate residue Asp$^{227}$. Below the alignment is the consensus sequence, represented as a sequence logo (Crooks et al., 2004), showing that Arg$^{220}$ is the most conserved residue in loop2.

monomer B are almost identical to those for monomer A.

In summary, this is the structure of a novel NHEJ polymerase complex (PolDom: NTP: metals: DNA), which represents an incomplete ternary complex (lacking a primer strand), ascribed as a pre-ternary complex. To our knowledge, this is the only structure of a polymerase-DNA complex containing a templated incoming base in the absence of a primer strand, a scenario that is most relevant for NHEJ polymerases. The biochemical and structural significance of this complex is described below.

## 6. Structural evidence supporting a pre-catalytic ternary complex.

We examined if any major structural rearrangements had occurred in the *Mt*-PolDom-DNA complex upon templated binding of UTP and metal ions in the active site. Notably, the polymerase in the pre-ternary complex adopts a similar overall conformation to that observed in the apo, nucleotide- and synaptic PolDom structures (Brissett et al., 2007; Pitcher et al., 2007), with only minor local conformational changes, as measured by pair-wise

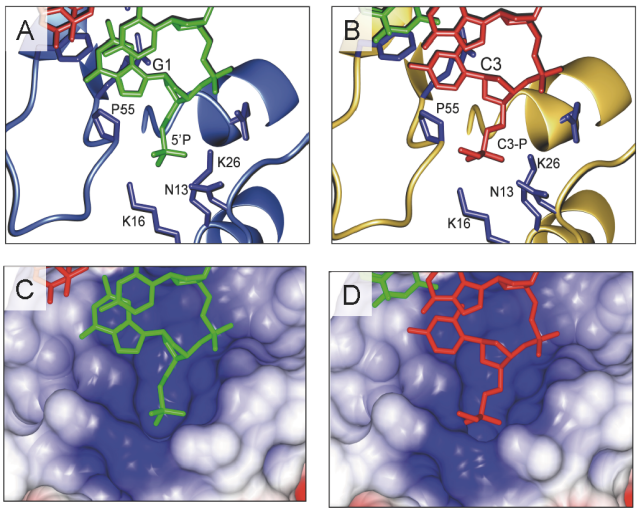



**Figure 61. Views of the 5'-phosphate binding pockets.** Binding pockets for monomer A (A&C) and monomer B (B&D). A&C) Stick and ribbon representations of the binding pocket with the conserved residues involved in forming the pocket highlighted in blue. Monomers A and B are coloured blue and gold, respectively. The templating/downstream DNA strands are coloured red/green (monomer A) or vice versa (monomer B). C&D) Solvent accessible surface representation of the same sites coloured to electrostatic potential.

alignment methods. The Root mean square deviation (RMSD) values are all well below 1.0, suggesting that there are no major global conformational differences between the various polymerase structures, although the current *Mt*-PolDom structure adopts a conformation most closely related to that found in the DNA binary complex. However, despite these overall similarities, a closer examination of the structural elements within the pre-ternary PolDom complex, revealed that a specific conserved loop (loop 2; residues 213-224) showed large RMSD values compared to the equivalent region in other *Mt*-PolDom structures (apo, GTP, dGTP and synaptic complex). Loop 2 exists as a $3_{10}$ helix in all of the previously determined structures. However, this helix unravels and adopts a random coil conformation in the pre-ternary complex (Fig. 60C). The significant conformational change of loop 2 results in Cα position shifts of ~6Å, inducing significant repositioning of the side-chains of residues on the loop. A comparison of the amino acid sequences of PolDom's revealed that loop 2 is conserved across the entire family of NHEJ primases (Fig. 60D), suggesting that this structural element could have a role in NHEJ. Closer inspection of the sequence of loop 2 identified an invariant arginine residue, $Arg^{220}$. Notably, the side-chain of $Arg^{220}$ significantly changed orientation, away from the active site, upon transition from the binary to the pre-ternary complex (Fig. 60C), suggesting that $Arg^{220}$ may play a specific role in catalysis. The functional significance of the movement of Loop 2 is described below.

A comparison of the active sites of the binary complex (PolDom bound to GTP and metal; {Pitcher et al., 2007) with the pre-ternary complex revealed significant difference in the positioning of the nucleotides in the active sites. Notably, the position of the triphosphate tail appears to dictate if a second metal ion, absent in the binary complex, can bind in the active site (Fig. 60A). The orientation of this tail is strongly influenced by template/nucleotide pairing. In the pre-ternary complex, the orientation of the templating adenine base (A6), forming a Watson-Crick base pair with the uridine base, is determined by a close stacking interaction with $Phe^{64}$ (Fig. 62A, right panel). Significantly, the incoming nucleotide base in the binary complex structure (with no templating base available) directly stacks against this phenylalanine (Fig. 60A) changing its overall location relative to that observed in the pre-ternary complex. Thus, $Phe^{64}$, invariant in NHEJ AEPs, plays a pivotal role in the sequential recognition and orientation of both the templating and incoming bases (Pitcher et al., 2007) and, therefore, also has an important role in determining the enzyme's affinity for a second active site metal ion.

Further support for the catalytic competency of the pre-ternary polymerase complex is provided by comparisons of the conformations of UTP (pre-ternary structure) with the nucleotides from the structures of a number of Pol X ternary complexes. For example, a comparison with the Polλ ternary complex (dUpnpp, PDB ID: 2PFO; Garcia-Diaz et al., 2007) reveals a pair-wise atom alignment of 0·66, suggesting that UTP (in the pre-ternary complex) is in a catalytically competent conformation (Fig. 60B). The lack of an incoming primer strand means that phosphotransfer is unable to occur and, therefore, the *Mt*-PolDom structure reported here must be defined as a pre-ternary catalytic complex. However, the conformation of the residues and cofactors (UTP and metals) in the active site suggests that the PolDom complex is optimally poised for extension upon engagement with an incoming primer strand, as will be described below.

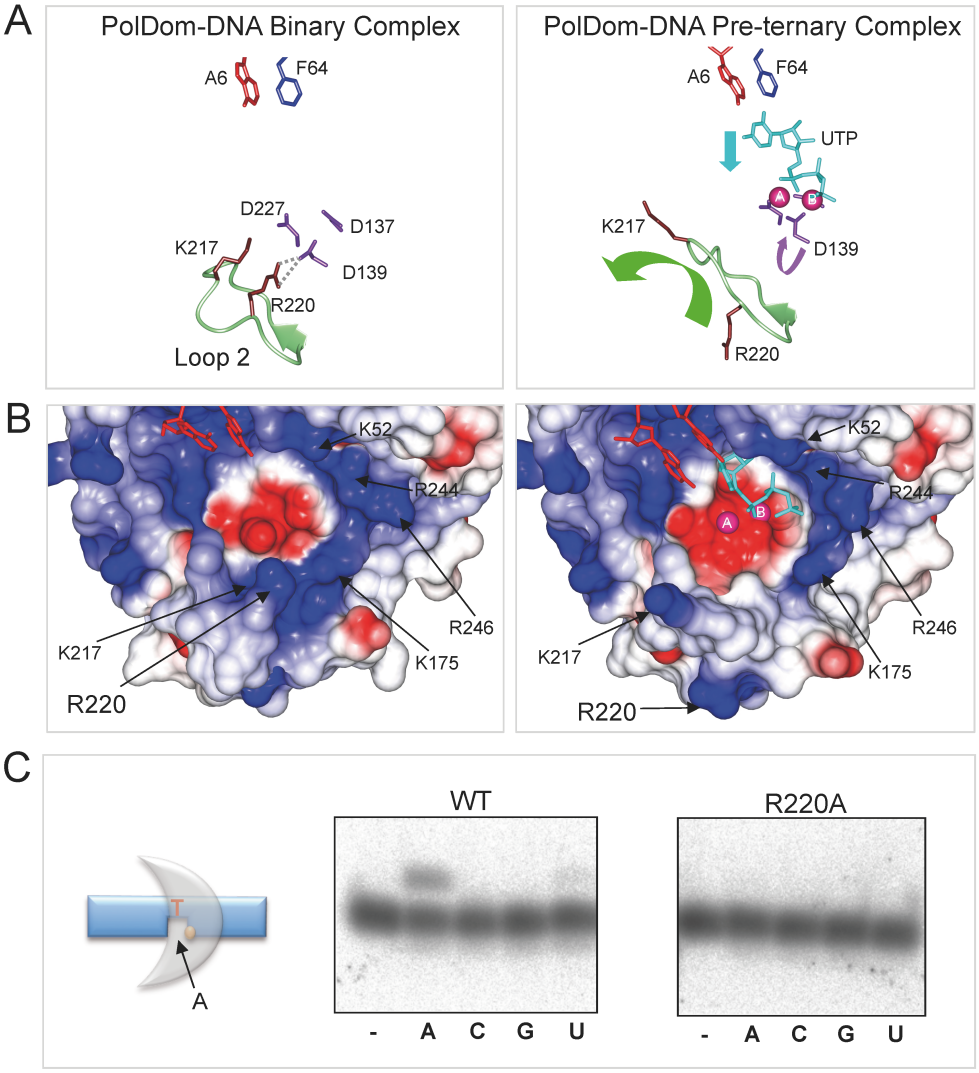



**Figure 62. Role of Loop 2 in Regulating the Active Site of PolDom**

A) Schematic view of the conformational changes affecting loop 2 (green) and some selected residues, comparing the Mt-PolDom:DNA binary complex (PDB ID: 2R9L; left) versus the preternary complex (right). Metal ions (A and B) are indicated as pink spheres. Incoming nucleotide (UTP) is colored cyan. The templating nucleotide (A6) present in the preternary complex, is indicated in red.

(B) Changes in electrostatic surface potential in the active site pocket, comparing te same structures as in A). Loop2 residues ($Lys^{217}$, $Arg^{220}$) are highlighted, as well as the residues involved in the formation of the positively charged ridge (in blue) above the active site.

(C) Gap-filling activity of Mt-PolDom loop 2 mutant, R220A. DNA synthesis reactions on a 5'-P-containing (yellow) 1 nt gapped substrate (5 nM) were carried out in the presence of either wild-type or mutant (R220A) Mt-PolDom (400 nM), 1 mM $MnCl_2$, and 100 mM of either the correct (ATP) or incorrect NTPs. After incubation for 30 min at 30°C, extension of the 5'-labeled primer strand was analyzed by 8 M urea-PAGE and autoradiography.

## 7. Repositioning of Loop 2 in the pre-ternary polymerase complex regulates metal binding at the active site.

Although there is a major reorientation of Loop 2, away from the active site, upon formation of the pre-ternary complex, what role does this loop transition play in the catalytic cycle of these NHEJ polymerases? Figure 62A&B show views of the active site of *Mt*-PolDom-DNA binary complex (left panels) and the pre-ternary PolDom complex (right panels) containing an incoming UTP base-paired to a templating adenine base (A6) with two metals ions also bound. A number of critical observations suggest that the pre-ternary complex is catalytically competent and that Loop 2 plays a critical role in activating the catalytic centre.

The first notable observation is that a salt-bridge formed between $Asp^{139}$ and $Arg^{220}$ (Loop 2), observed in the DNA binary complex (Fig. 62A, left), is disrupted in the pre-ternary complex (Fig. 62A, right). This causes a rotation of the arginine side-chain (~180°) away from the active site, enabling $Asp^{139}$ to adopt a conformation where it can favorably act as a ligand for the second catalytic metal ion (A), not present in the catalytically inactive structures (Pitcher et al., 2007). Accordingly, mutation of this arginine to alanine (R220A) significantly reduced *Mt*-PolDom´s ability to incorporate nucleotides (Fig. 62C) but, significantly, did not alter its affinity for gapped-DNA substrates (data not shown). It appears that binding of this second metal ion itself triggers this conformational change and this significant rearrangement at the active site implies that $Arg^{220}$ may play a direct role in catalysis.

Based on both these structural and biochemical observations, we propose that Watson-Crick base pairing of an incoming nucleotide (forming a chelate with metal B), complementary to the templating base

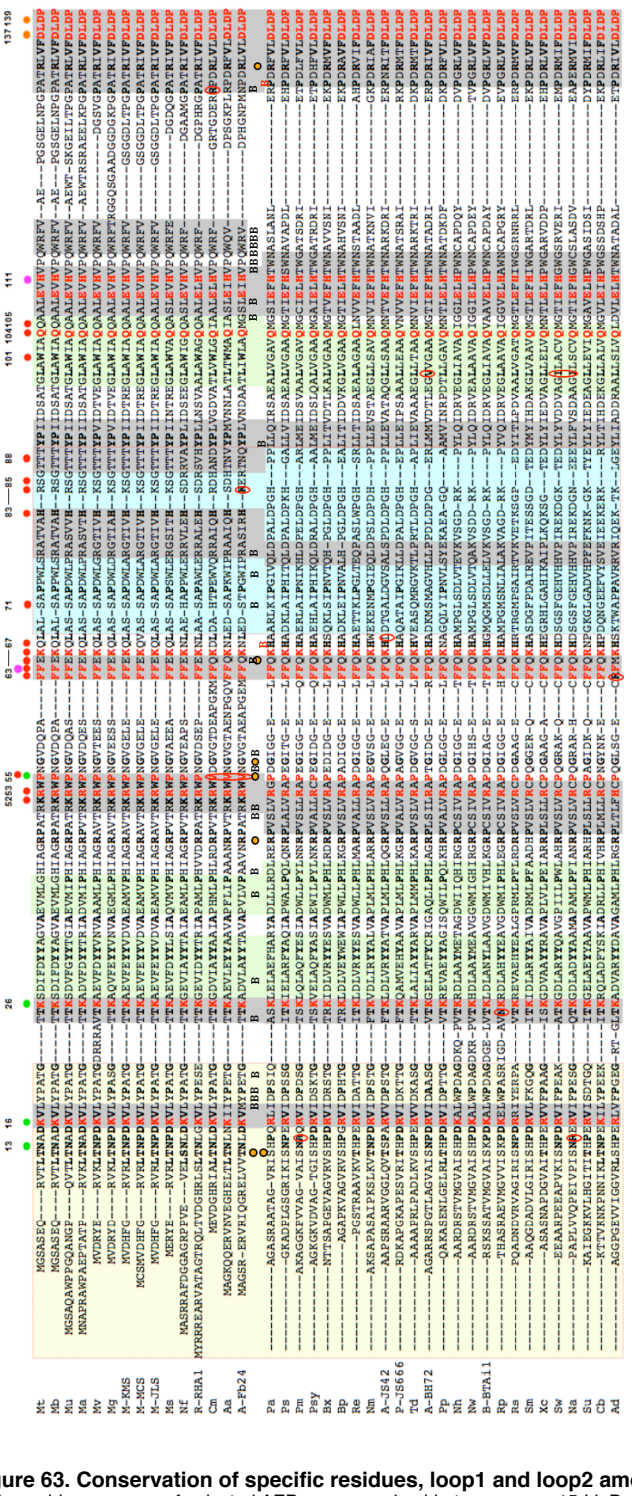



**Figure 63. Conservation of specific residues, loop1 and loop2 among LigD members.**
Amino acid sequences of selected AEPs are organized in two groups: 15 LigD sequences having a *Mt*-LigD like architecture, in which the polymerization domain is at the N-terminus, all of them being Actinobacteria (Mt, *Mycobacterium tuberculosis*; Mb, *Mycobacterium bovis*; Mu, *Mycobacterium ulcerans*; Ma, *Mycobacterium avium*; Mv, *Mycobacterium vanbaalenii*; Mg, *Mycobacterium gilvum*; M-KMS, *Mycobacterium sp. KMS*; M-MCS, *Mycobacterium sp.MCS*; M-JLS, *Mycobacterium sp. JLS*; Ms, *Mycobacterium smegmatis*; Nf, *Nocardia farcinica*; R-RHA1, *Rhodococcus sp. RHA1*; Cm, *Clavibacter michiganensis*; Aa, *Arthrobacter aurescens*; A-FB24, *Arthrobacter sp. FB24*); 25 LigD sequences having a *Pa*-LigD like architecture, in which the polymerization domain is at the C-terminus, 24 of them being Proteobacteria (Pa, *Pseudomonas aeruginosa*; Ps, *Pseudomonas stutzeri*; Pm, *Pseudomonas mendocina*; Psy, *Pseudomonas syringae*; Bx, *Burkholderia xenovorans*; Bp, *Burkholderia pseudomallei*; Re, *Ralstonia eutropha*; Nm, *Nitrosospira multiformis*; A-JS42, *Acidovorax sp. JS42*; P-JS666, *Polaromonas sp. JS666*; Td, *Thiobacillus denitrificans*; A-BH72, *Azoarcus sp. BH72*; Pp, *Pseudomonas putida*; Nh, *Nitrobacter hamburgensis*; Nw, *Nitrobacter winogradskyi*; B-Btai1, *Bradyrhizobium sp. BTAi1*; Rp, *Rhodopseudomonas palustris*; Rs, *Rhodobacter sphaeroides*; Sm, *Sinorhizobium medicae*; Xc, *Xanthomonas campestris*; Sw, *Sphingomonas wittichii*; Na, *Novosphingobium aromaticivorans*; Cb, *Coxiella burnetii*; Ad, *Anaeromyxobacter dehalogenans*), and 1 sequence of Acidobacteria (Su, *Solibacter usitatus*); The portion aligned corresponds to residues of the polymerization domain (PolDom) in each of the two LigD subfamily members. Boxes indicate secondary structure elements (beta strands in gray; alpha-helices in green) according to the crystal structures of *Pa*-LigD complexed with ATP (PDB ID: 2faq), *Mt*-LigD either complexed with GTP (PDB ID: 2irx) or with a NHEJ DNA intermediate, and *Pfu*-AEP (PDB ID: 1G71). The areas indicated with light blue color correspond to the "loop1" and "loop2" regions of *Mt*-LigD PolDom. The region indicated with a light yellow area contains the *Mt*-LigD Lys[16] residue, shown to interact with the 5'-phosphate at a DNA end. Dots above the *Mt*-LigD sequence indicate residues involved in interacting with the 5'-phosphate (Asn[13], Lys[16], Lys[26], Pro[55]; green dots), with nucleotides at both strands of a T/D intermediate (Lys[52], Arg[53], Pro[55], Phe[63], Phe[64], Glu[65], Lys[66], Gln[67], Ser[71], H83, Arg[84], Ser[85], Thr[88], Trp[101], Gln[104], Gln[105], Lys[217]; red dots), or involved in catalysis (Asp[137], Asp[139], Arg[220] and Asp[227]; orange dots). Pink dots above both *Mt*-LigD and *Pa*-LigD sequences indicate residues involved in interaction with the incoming NTP (Phe[64] and His[111], Ser[172], Ser[174], Lys[175], Gly[176], His[178], Thr[236], Ile[238], Arg[244] and Arg[246]). Amino acid residue colouring reflects the degree of conservation of this N-terminal portion of the polymerization domain between the subfamilies (red: invariant or almost invariant; bold: conservative residues in most of the sequences.

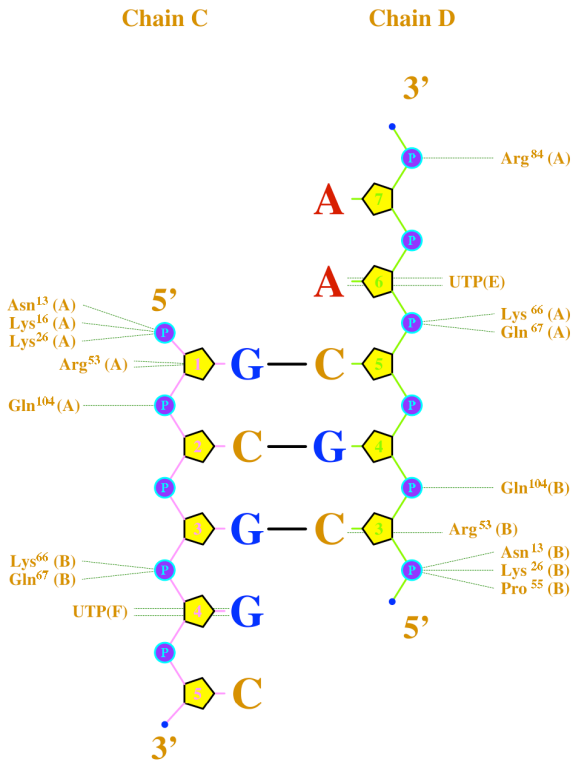



**Figure 64. Interactions between *Mt*-PolDom and the 3'-overhanging DNA duplex.**
The complex is formed by monomers A & B of PolDom and DNA chains C and D. Green arrows, indicate hydrogen bonds; P, DNA backbone phosphates; DNA residue numbers are shown inside the pentagons representing ribose sugars.

nearest to the recessive 5'-P, acts as the primary conformational switch that makes the polymerase active site catalytically competent. Thus, after selection of the correct nucleotide, a second metal (A) is recruited to the active site, perhaps still in a non-competent location requiring the interaction between $Arg^{220}$ and $Asp^{139}$ to provide the necessary space/charge to bind this second metal ion. Subsequently, a metal-driven conformational change of loop 2 leads to the breakage of the $Arg^{220}$- $Asp^{139}$ salt bridge, enabling $Asp^{139}$ to act as a metal-binding ligand to promote repositioning of metal A into a catalytically competent location within the active site. Therefore, the capacity to form a salt bridge between $Arg^{220}$ and $Asp^{139}$ (abolished in the R220A mutant), which keeps the active site in a "stand-by" mode, is an important requisite step preceding catalysis. That pre-activated complex, loaded with a complementary nucleotide can occur at a single DNA end, before a second 3'-protruding end provides an attacking primer for nucleotide insertion and NHEJ.

## 8. Formation of a pre-ternary complex provides a catalytic advantage during NHEJ.

### 8.1 Fluorescence analysis of the pre-ternary complex.

To support the crystallographic model for a pre-ternary polymerase complex, we measured the formation and stability of PolDom pre-ternary complexes in solution, adopting established fluorescent methodologies used for analyzing DNA polymerases ternary complexes (Fidalgo da Silva et al., 2002; Kumar et al., 2008; Zhang et al., 2007). Formation of a pre-ternary complex was assessed by adding UTP to a preformed fluorescent complex of *Mt*-PolDom:DNA:$Mn^{2+}$, in which the 3'-protruding DNA end contains 2-aminopurine (2AP) as the templating base. We observed that a stable pre-ternary PolDom complex was formed in solution, as pairing of UTP to the templating base quenched the intrinsic fluorescent signal of 2AP (Fig. 65A). Titration of the UTP concentration enabled the determination of a binding constant ($K_d$) for nucleotide in the pre-ternary complex (PolDom:DNA:UTP:$Mn^{2+}$). Figure 65A shows the normalized fluorescence of the preformed *Mt*-PolDom:DNA:$Mn^{2+}$ complex with varying concentrations of UTP and the ensuing quenching that accompanies base pairing of 2-aminopurine and UTP. From these data we calculated a $K_d$ of 0.646 mM for the binding of UTP to 2-aminopurine, determined by fitting relative quenching values to a ligand binding equation with a non-specific binding factor. The sub-micromolar range $K_d$ for UTP binding closely correlates with the binding constants determined for nucleotide binding for *Mt*-PolDom (Pitcher et al., 2007) and other DNA polymerase ternary complexes (Kumar et al., 2008).

### 8.2 Formation of the pre-ternary complex in solution assessed by EMSA.

Next, we evaluated the formation and stability of both binary and pre-ternary DNA complexes in solution by EMSA using a T/D DNA molecule with a short 3´ overhang (GTT-3´), in the presence/absence of activating metals and incoming nucleotides. As shown in figure 65B, in the absence of metal ions, *Mt*-PolDom produced a stable binary

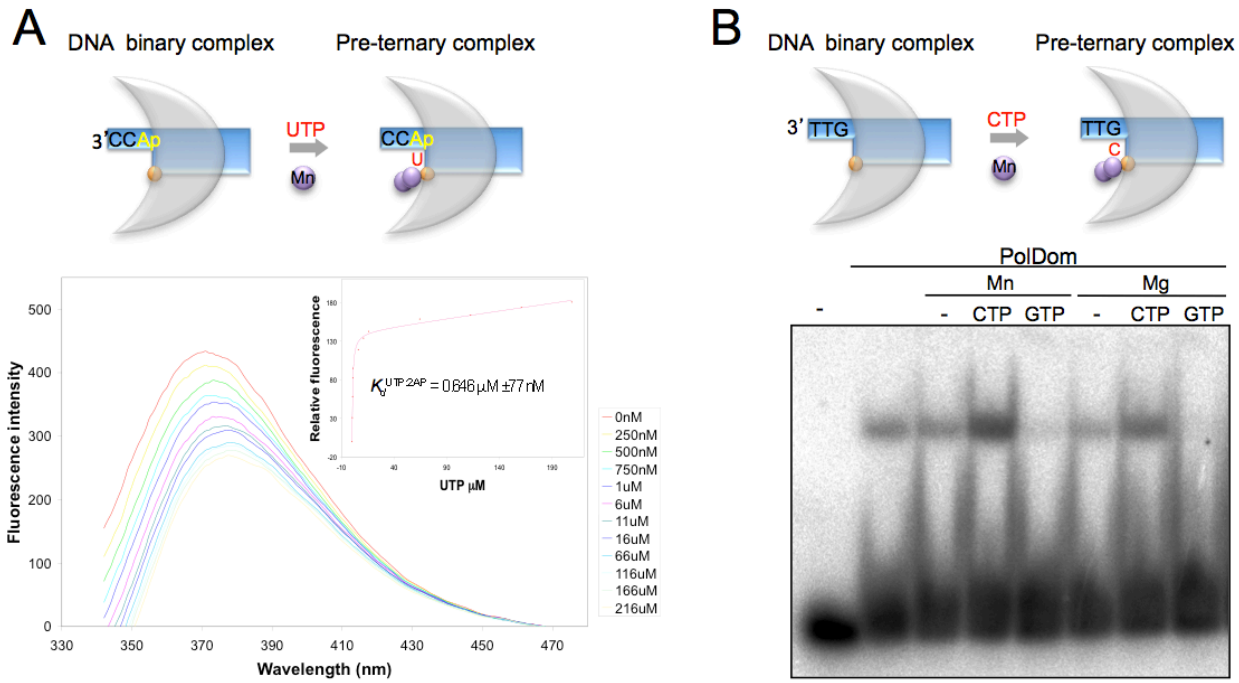



**Figure 65. Formation of a Stable Preternary Polymerase Complex in Solution**
A) Titration of Mt-PolDom:2AP-labeled DNA complex with free UTP in the presence of MnCl2. The main graph shows the normalized fluorescence of Mt-PolDom:2AP-labeled DNA complex (200 nM Mt-PolDom, 200 nM ddC terminated 2-aminopurine labeled DNA, 50 mM Tris [pH 7.5], 150 mM NaCl, 10 mM MnCl2, and 4% glycerol) titrated with varying concentrations of UTP. Inset is a graph showing the relative fluorescence quenching as a function of UTP concentration (blue diamonds) overlaid with data fit to the nonlinear regression ligand binding equation (pink line). B) A stable preternary complex formed in solution. As detected by electrophoretic mobility shift assays (EMSA), Mt-PolDom binds 3'-protruding (GTT-3') T/D ends with a 5'-P to form a PolDom:DNA binary complex (arrow). Addition of activating metal ions and a complementary nucleotide (CTP) allows formation of a more stable preternary complex. Conversely, a stable preternary complex is not formed in the presence of a noncomplementary (GTP) nucleotide. EMSA assays were performed with 600 nM Mt-PolDom and 5 nM DNA. When indicated, cofactors (100 $\mu$M MgCl2, 10 $\mu$M MnCl2,) and nucleotides (either CTP or GTP 100 $\mu$M) were added.

complex with DNA, observed as a sharp retarded band. Addition of either 10 $\mu$M $Mn^{+2}$ or 100 $\mu$M $Mg^{+2}$ did not significantly affect the formation and stability of such a binary complex. However, addition of both activating metal ions and CTP (the required nucleotide to form a correct pre-ternary complex) significantly enhanced the stability of the complex formed (a pre-ternary complex). Notably, addition of a non-complementary nucleotide (GTP) had a negative effect on the stability of the DNA binary complex.

### 8.3 Evaluation of catalysis after formation of the pre-ternary complex.

Although PolDom can form a stable pre-ternary complex in solution, is this complex catalytically active? Moreover, is such a preformed complex beneficial for catalysis? To address these questions, we tested if formation of a pre-ternary complex, prior to the encounter with a primer resulted in altered primer extension activity. For that, we used a trinucleotide to minimize terminal transferase-like nucleotide additions driven by *Mt*-PolDom. A scheme of the experiment is shown in figure 66A. Figure 66B shows that the amount of 3G primer extended (dG-dG-dG-CMP) is significantly higher (~30% more efficient) when PolDom was pre-incubated with the templating molecule (..GCC3´) + correct nucleotide (CTP) + metal ions ($MnCl_2$) in the absence of primer, than when all reactants were added simultaneously. That beneficial effect of the pre-incubation did not occur when any other combination of co-factors were pre-incubated (Fig. 66B). Therefore, the only case in which a catalytic advantage was obtained (see the reaction kinetics at Fig. 66C) includes the four components needed to form a pre-ternary complex. Similar results were also obtained with primers of different lengths and matching different DNA ends (not shown). These data support the catalytic significance of a pre-ternary complex in solution and suggest that such complexes enable

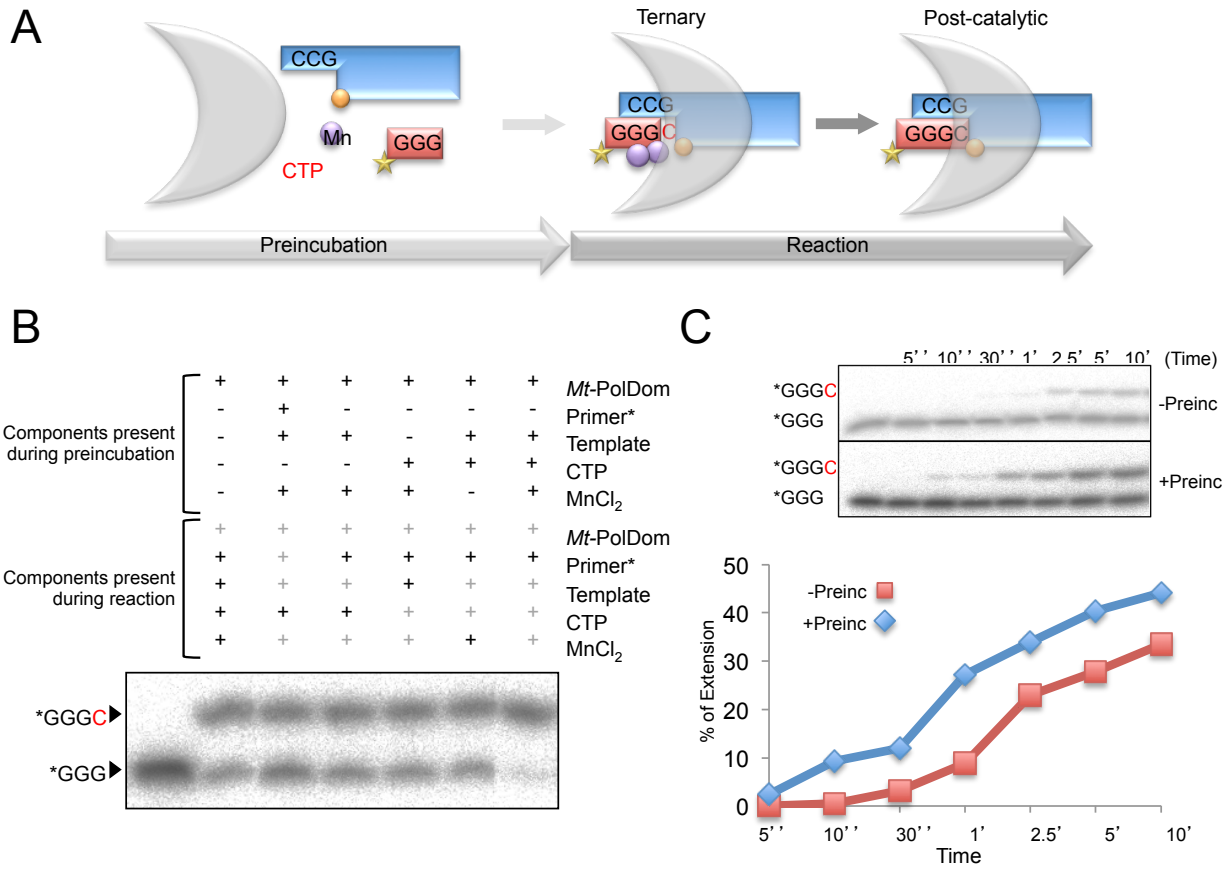



**Figure 66. Formation of a Preternary Complex Enhances Nucleotide Extension off an Incoming Primer Strand**
A) Schematic representation of the experimental steps reported in parts (B) and (C) (below). Some of the components needed for the extension reaction (Mt-PolDom [gray crescent], a template molecule [blue] with a 50 P [yellow dot] and a GCC-30 protrusion, dGdGdG [red] as primer, CTP, and manganese ions) were allowed to interact during a preincubation step, as indicated in (B) for each case. Subsequently, the remaining components required for the reaction to occur were added, and the extension step can proceed. B) NHEJ polymerase extension reactions, following the scheme described in (A). Concentrations used were 600 nM Mt-PolDom, 1mM MnCl2, 100 mM CTP, 5 nM labeled primer, cold CCG30, and 25 nM template DNA. During preincubation for 30 min at 4ºC, the components present are indicated by a + sign, while missing components are indicated by a – sign. At the start of the reaction period, added components are indicated by a black + sign, while components already present are indicated by a gray + sign. Reactions were performed for 30 min at 30ºC, and the samples were run on a 6 M Urea-20% polyacrylamide gel and autoradio- graphed. C) Kinetics of NHEJ polymerase extension reactions following the scheme described in (A). –Pre-inc indicates preincubation of Mt-PolDom only; +Pre-inc indi- cates preincubation of the components needed to form the preternary complex (Mt-PolDom, nucleotide, template DNA, Mn2+). Insertion reactions were stopped at the times indicated, and samples loaded on a 6 M Urea-20% polyacrylamide gel and autoradiographed.

these enzymes to utilize additional activities (e.g. terminal transferase) to extend off otherwise unstable termini (e.g. very short /blunt termini).

## 9. Functional significance of the pre-ternary complex.

In the pre-ternary complex, the position of a number of conserved positively charged residues (e.g. Lys$^{175}$ & Arg$^{246}$), which line the active site pocket, has altered significantly with respect to the PolDom-DNA binary complex, modulating the shape and charge of the ridge (Fig. 62B). This conformational change enlarges the cavity to facilitate access to an incoming primer strand (Fig. 62B, right panel), in agreement with the catalytic advantage provided by a pre-formed pre-ternary complex. Replicative AEPs (having an RNA primase function) are able to bind a "primer" as small as a single NTP, allowing the synthesis of a dinucleotide at the initiation step of RNA primer synthesis (Fig. 67A). As the bacterial NHEJ polymerases are closely related to these canonical AEPs, we tested the minimal length of DNA (incoming end) that can be accepted as a "primer" by the pre-ternary PolDom complex.
As shown in Figure 67B, three different molecules were tested as potential primers to be extended with CTP (the nucleotide pre-loaded in a pre-ternary complex formed at a 3'-protruding [CCG] DNA end). The scheme at the left shows the degree of complementarity of the three primers used (3G, 2G or G). Primer extension (p + CMP) was observed with 3G and 2G primers, but not with GTP, and was mostly dependent on the presence of the DNA end (pre-ternary complex). Figure 67C shows that the minimal primer (2G) could only be extended when

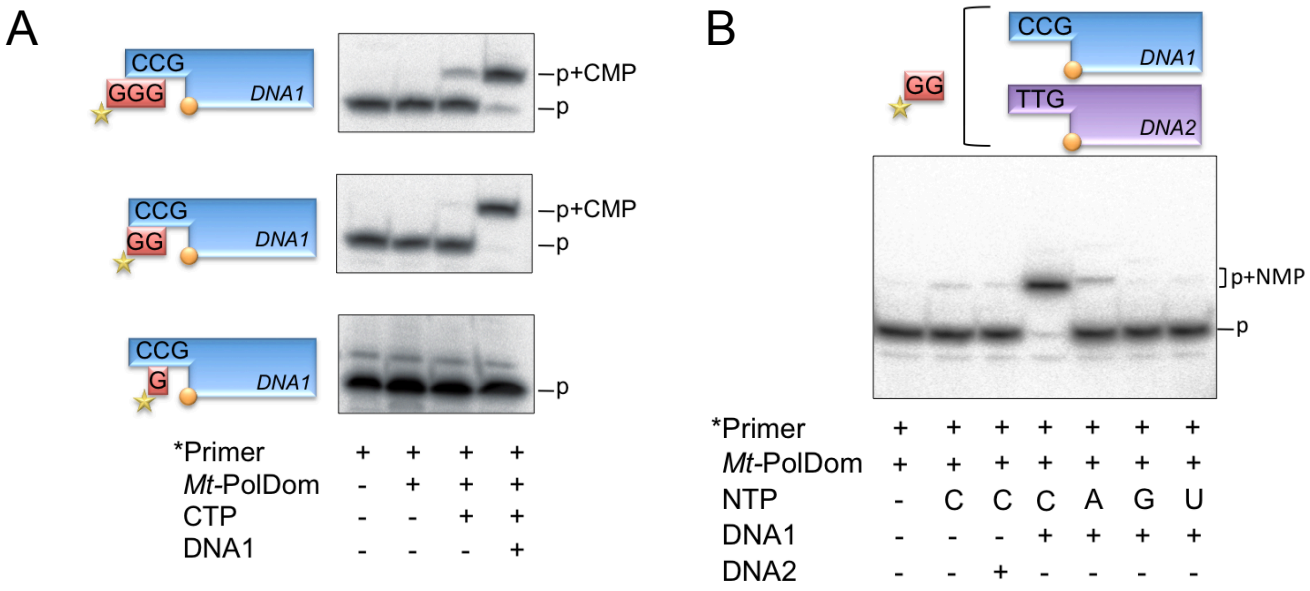



**Figure 67. Minimal Length of DNA that Can Be Accepted as a "Primer" by the Preternary PolDom Complex**
A) A 30-protruding (GCC-30) DNA molecule (DNA1; 25 nM), with a 5'-P (yellow sphere) was used as the DNA end to form a preternary complex in the presence of 600 nM Mt-PolDom, 1 mM MnCl2, and 100 mM CTP. Different DNA molecules (dGdGdG, dGdG, or dG; indicated in red), labeled at the 50 end, were tested as primer strands (5 nM) to accept the CTP substrate preloaded in the preternary complex. After incubation for 24 hr at 4ºC, primer extension products (p+CMP) were separated by 8M Urea-30% PAGE and detected by autoradiography. B) As in (A), but with two different DNA molecules (DNA1 and DNA2), both capable of forming a pre-ternary complex

the DNA terminus was complementary (CCG), and not with a non-complementary end (TTG). Moreover, a complementary NTP (a requisite to form and stabilize a pre-ternary complex; see Fig. 62B) was required to form an extended product (dG-dG-CMP; Fig. 67C).

**10. Mt-PolDom-mediated NHEJ synapsis.**

Although Mt-PolDom possesses a mild terminal-transferase like activity, this enzyme is preferentially a DNA template-directed polymerase (ref). As shown in figure 68A (left panel), by using a single-stranded homopolymer (poly-dA) as the sole DNA, preferential extension with a single nucleotide (UTP) occurred but the reaction was very inefficient. Conversely, when another homopolymer (poly-dT) was used, the nucleotide preferentially inserted was now ATP (not shown), again complementary to the homopolymer. This suggests that the nucleotide incorporated into a poly-dA molecule did not occur *via* terminal transferase but directed by a second DNA molecule (a NHEJ reaction), perhaps favored by a dimeric arrangement as depicted in the figure, promoted by the specific and strong interaction of each PolDom monomer with a 5´-P residue (Brissett et al., 2007). The inefficiency of this reaction could be the consequence of a non-complementary synapsis, in which the primer-terminus would be unpaired in the vicinity of the NTP binding site. In contrast, by adding a 5-fold excess of an unlabeled T/D molecule with a 3´ protrusion of 4 nt (GTTT-3´), the nucleotide preferentially inserted into the labelled poly-dA was CTP, complementary to the dG residue neighbour to the 5'-P of the T/D molecule (Fig. 68A, middle panel). In this case, both synapsis and precise approach of the 3´-terminus of the poly-dA to the active site is facilitated by the complementarity of the last three 3´-terminal bases of each molecule (AAA-3'/TTT-3'). Strikingly, by adding an even higher amount of the T/D (GTTT-3´) molecule (~10-fold excess over poly-dA; Fig. 68A, right panel), extension of poly-dA with CTP was significantly inhibited, suggesting that a stable dimeric arrangement of two T/D molecules (non-labelled) can occur, even though their protrusions are not complementary (GTTT-3'/GTTT-3'), similarly to what was observed in the structure of the Mt-PolDom dimeric arrangement with two DNA ends forming an imperfect synapsis (Brissett et al., 2007). However, that kind of imperfect but stable synapsis was catalytically incompetent, as it was demonstrated on labeled T/D (GTTT-3´) molecules, by addition of activating metals and each of the four NTPs (Fig.

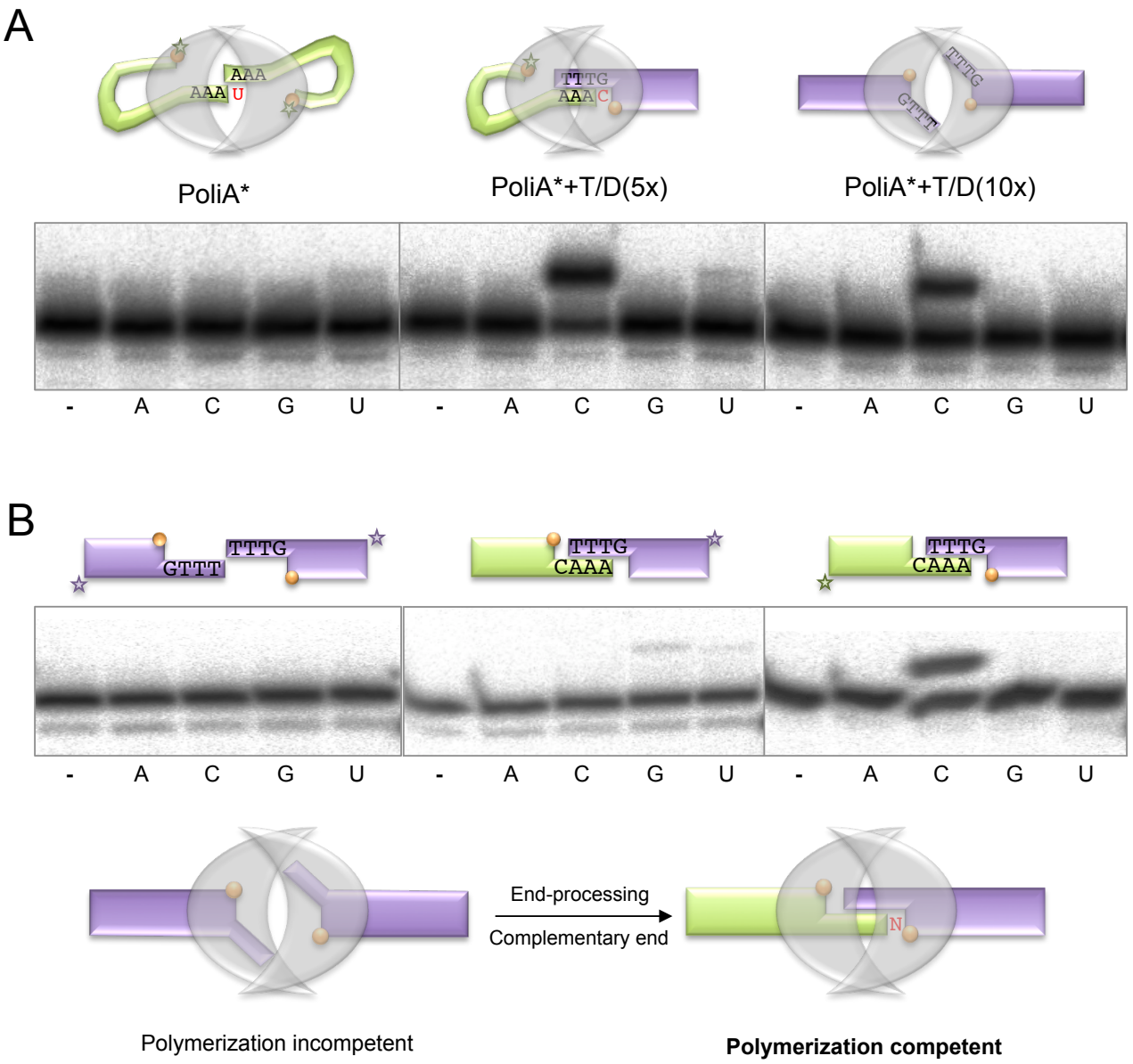



**Figure 68. PolDom-mediated NHEJ synapsis: dimer formation.**
A) NHEJ reactions were performed with 600 nM PolDom using various substrates: a homopolymeric single-stranded substrate (polydA) and a 3'-protruding substrate formed with the oligonucleotides 1TTTG and 1D-NHEJ. The orange balls indicate the presence of a 5'-P group in the substrate. When indicated, each of the four NTPs (100 $\mu$M) were added in the presence of 1 mM $MnCl_2$. The star shows the position of the radioactive label. B) NHEJ reactions were performed with 600 nM PolDom using various substrates formed with the oligonucleotides 1TTTG or 1AAAC with 1D-NHEJ. The orange balls indicate the presence of a 5'-P group in the substrate. The star shows the position of the radioactive label. When indicated, each of the four NTPs (100 $\mu$M) were added in the presence of 1 mM $MnCl_2$.

68B, left panel): no incorporation was observed in any case. Conversely, when two compatible ends (GTTT-3'/CAAA-3') were simultaneously present, each 3'-protruding end could be preferentially extended with the nucleotide complementary to the templating base neighbour to the 5'-P, provided in trans by the opposite end (Fig. 68B, middle and right panels), as expected from bona-fide NHEJ reactions. Similar results were obtained with T/D molecules having shorter protrusions (see Fig. 69A for incompatible ends, and Fig. 69B&C for compatible ends). Therefore, as depicted in Fig, 68B, it can be inferred that a highly stable complex is formed even when the two 3'-protruding ends are not complementary and hinder catalysis, perhaps to allow/favor further nucleolytic processing by the nuclease domain in the same LigD enzyme in order to become catalytically competent.

As a new result from the ongoing collaboration with Prof. Aidan Doherty, a very recently solved crystal structure is shown in figure 70A in frontal view: the complex contains two Mt-PolDom monomers (blue

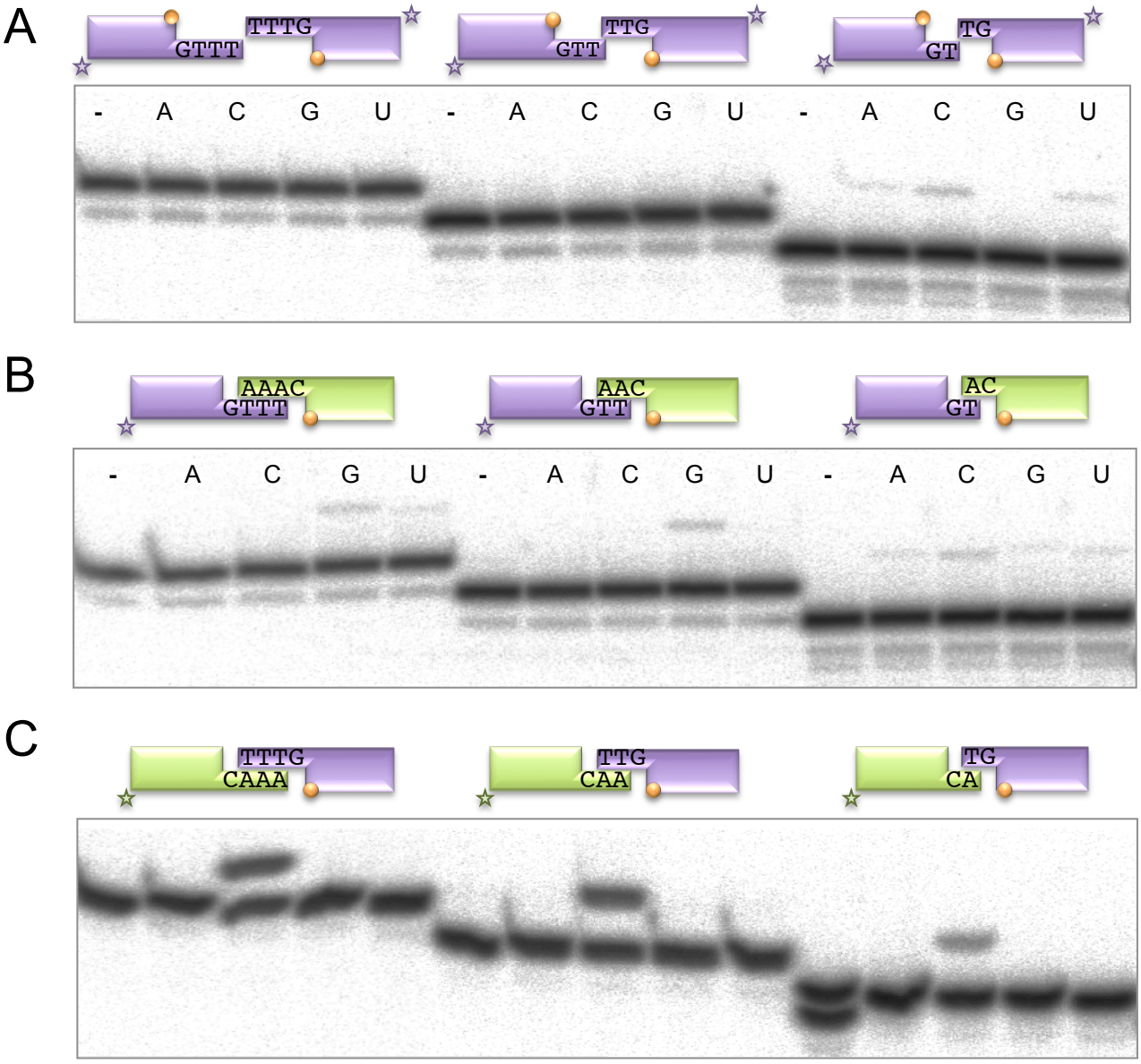



**Figure 69. PolDom-mediated NHEJ synapsis: importance of the length of the 3'-protrusions in the connection.**
A) NHEJ reactions were performed with 600 nM PolDom using various 3'-protruding substrates, formed with the oligonucleotides 1TTTG, 1TTG or 1TG with 1D-NHEJ. The orange balls indicate the presence of a 5'-P group in the substrate. The star shows the position of the radioactive label. When indicated, each of the four NTPs (100 $\mu$M) were added in the presence of 1 mM $MnCl_2$. B&C) NHEJ reactions were performed with 600 nM PolDom using various substrates formed with the oligonucleotides 1TTTG, 1TTG, 1TG, 1AAAC, 1AAC or 1AC with 1D-NHEJ. The orange balls indicate the presence of a 5'-P group in the substrate. The star shows the position of the radioactive label. When indicated, each of the four NTPs (100 $\mu$M) were added in the presence of 1 mM $MnCl_2$.

and yellow) bridging two compatible DNA ends (orange and purple). This new synaptic structure brings light into the orchestration of a dimer in the context of very short 3' protrusions, specifically addressing the issues of correct primer binding and templating base selectivity. In figure 70B a top view of the complex can be observed: the predicted role of loop 1 in maintaining the synapsis (Brissett et al., 2007) is confirmed here and loop 2 seems to be also playing a role in the connection.

## 11. Formation of a functional NHEJ complex: dimer *versus* monomer.

Once we demonstrated a catalytically competent synapsis of two DNA ends by PolDom, a more in depth analysis of the mechanism of NHEJ reaction with the four players involved (polymerase, incoming nucleotide, and the two 3'-protruding DNA ends) was necessary. Thus, the first question we wanted to

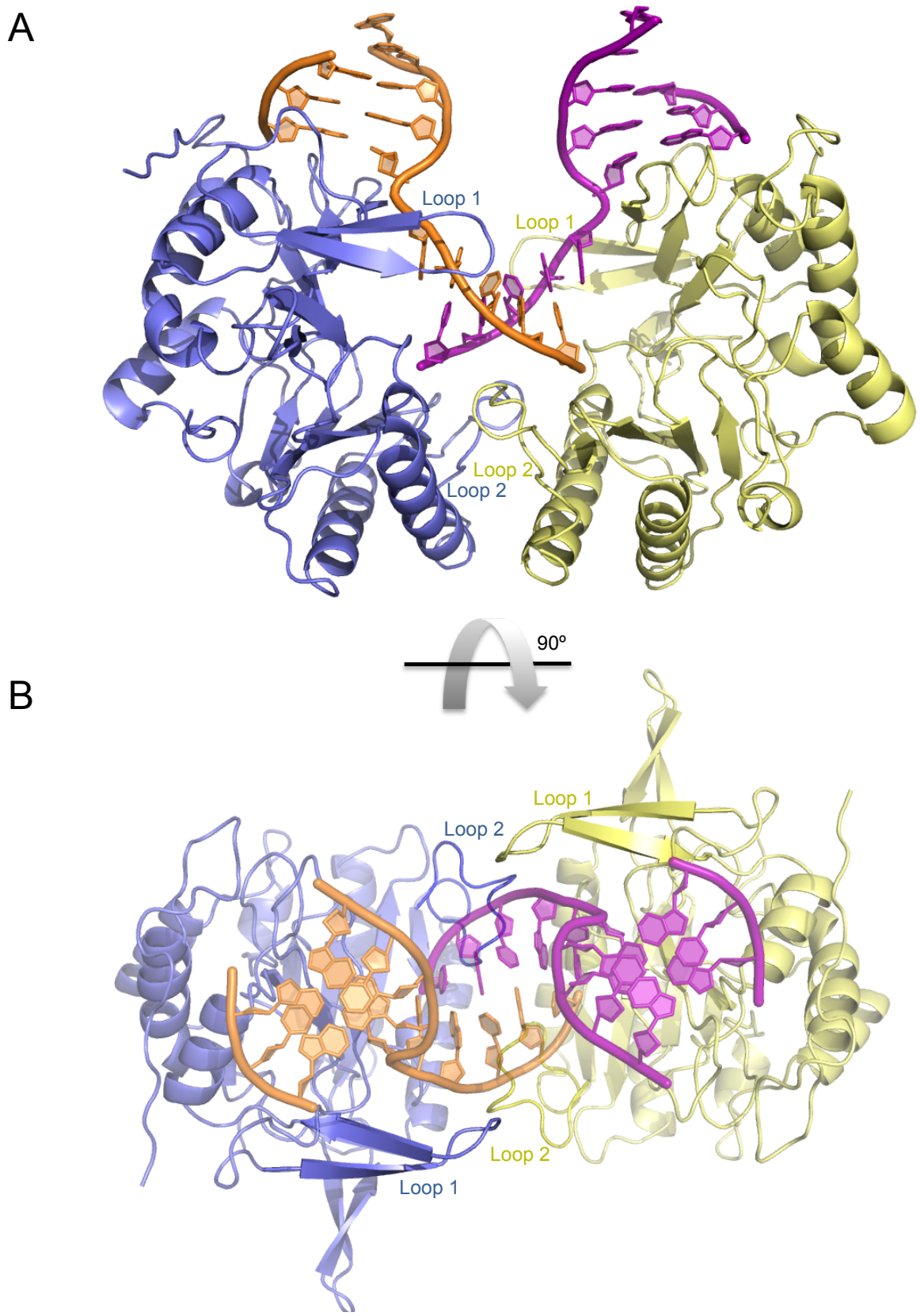



**Figure 70. Fully-complementary synapsis of two DNA ends mediated by a PolDom dimer.**
A) Frontal view of the new crystal structure showing two monomers of PolDom holding two pieces of DNA. The two monomers are colore yellow and blue, the two DNA substrates are shown in purple and orange. Loop1 and loop2 are indicated in the figure. B) Top view of the same complex.

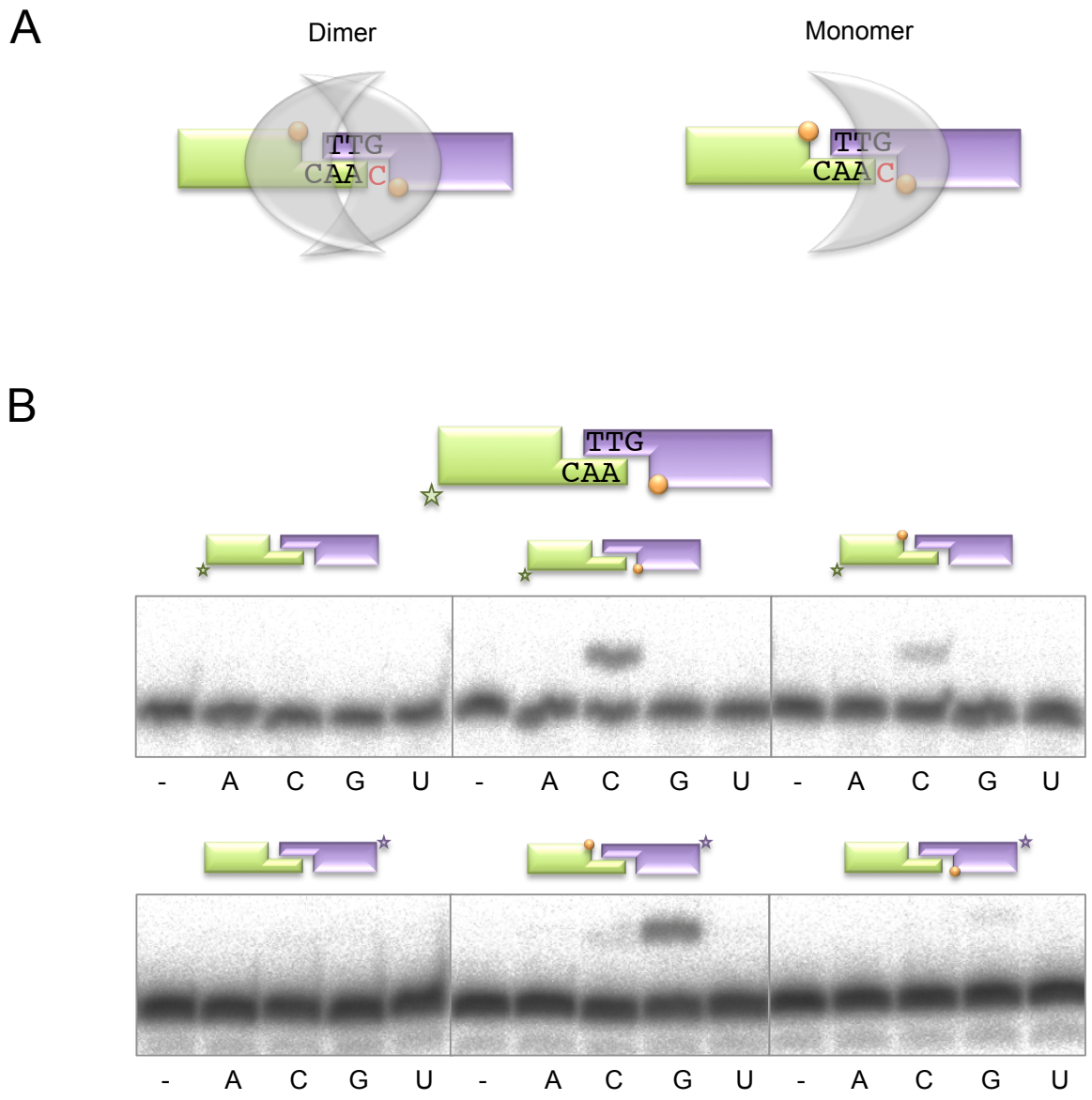



**Figure 71. Distiguishing between monomer and dimer: importance of the 5'-P group.**
A) Schematic representation of a PolDom dimer and a monomer. The protein is shown as a grey surface, the two DNA substrates are shown in green and mauve, orange balls show the position of the 5'-P groups, the incoming nucleotide is shown in red. B) NHEJ reactions were performed with 600 nM PolDom using various substrates formed with the oligonucleotides 1TTG or 1AAC with 1D-NHEJ. The orange balls indicate the presence of a 5'-P group in the substrate. The star shows the position of the radioactive label. When indicated, each of the four NTPs (100 $\mu$M) were added in the presence of 1 mM $MnCl_2$.

address was whether formation of a Mt-PolDom dimer (as occurred either during the imperfect synapsis previously described (Brissett et al., 2007), or in the new crystal of a perfect synapsis described here) was an absolute requirement for NHEJ of short complementary ends, or the reaction could be performed by a single PolDom monomer (Fig. 71A). For this, we used a set of 3'-protruding molecules having either none, or one 5'-P at only one of the two DNA ends (See schemes in Fig. 71B). As expected, the reaction was specific (template-directed) and required at least one phosphate group at the 5'-end (no reaction was obtained with a 5'-OH group at both ends; Fig. 71B). When the 5'-P group was in the GTT-3' end (mauve) there was a relevant insertion of CMP into the opposite 3'-protruding end, that can be detected when the latter is labelled. Conversely, when the 5'-P was present in the CAA-3' end (green), GMP was selectively and efficiently inserted at the other end. In summary, in a 2nt base-paired synapsis, formation of a Mt-PolDom dimer (that would require a 5'-P group at each DNA end) is not strictly required. Probably, two complementary ends can configure a "perfect gap" that can be recognized/stabilized by a single polymerase monomer. However, there was also a

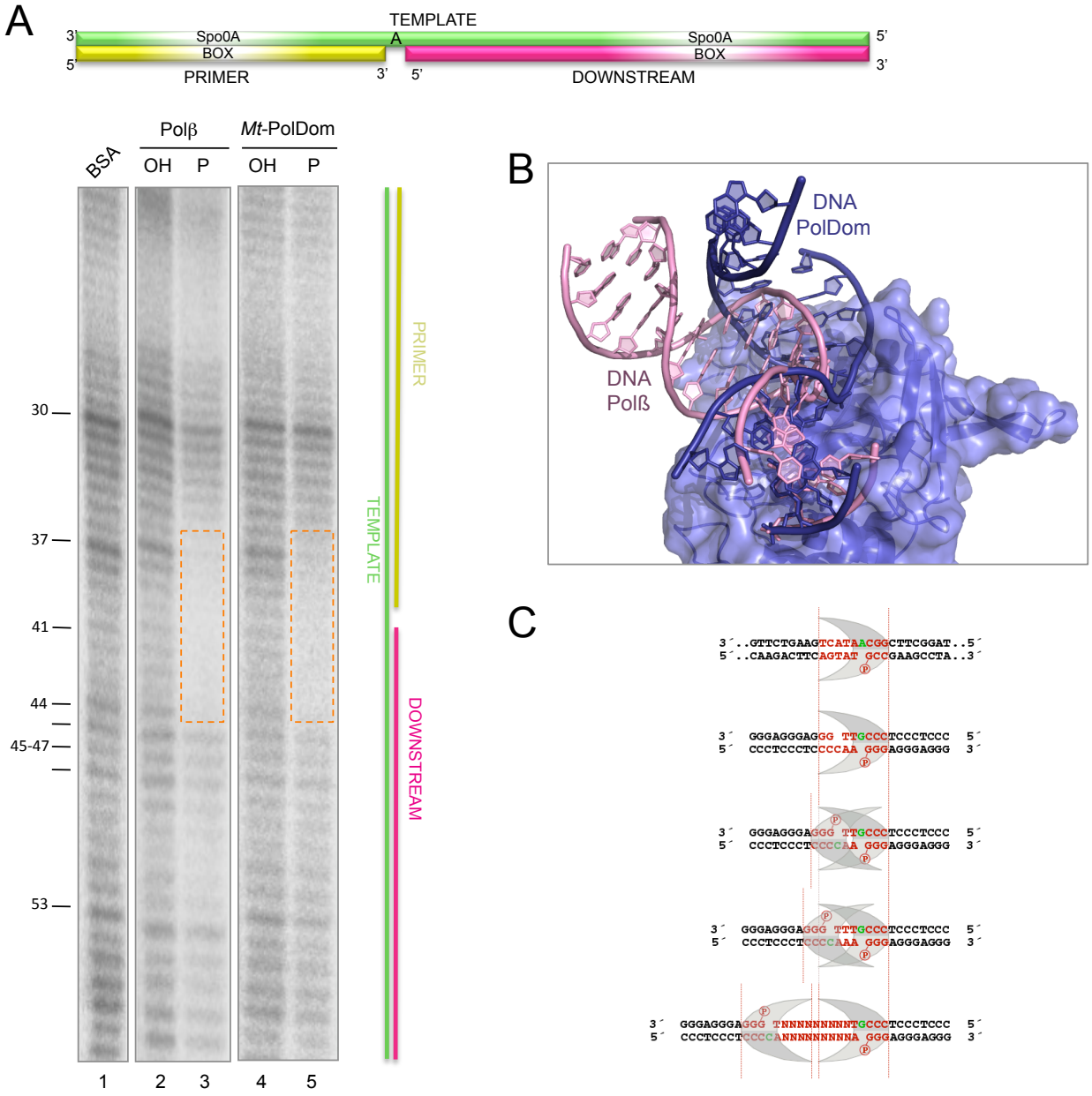



**Figure 72. Evidence of the monomeric form of PolDom on a gapped substrate.**
A) Scheme of the substrates used for the footprinting assays. To produce this substrate the oligonucleotides FP-T (template), FP-P (primer) and FP-D (downstream) were hibridized. B) Footprinting assay of Polß (5 $\mu$g) and PolDom (5 $\mu$g) was conducted as described in Materials and Methods. 10 $\mu$g of BSA were added to the control lane. D) Views of the crystal structure of the PolDom complementary synapsis, in which the protein is shown in semi-transparent blue surface, PolDom's own DNA substrate is shown in dark blue sticks, and the superimposed Polß DNA is shown in pink. C) Schematic representation of a monomeric and a dimeric arrangement of PolDom bridging two DNA ends and the number of nucleotides protected (shown in red). The tempating base is shown in green.

significant NHEJ reaction when the only 5'-P group was at the labelled DNA end. Therefore, it cannot be discarded that a binary or pre-ternary complex formed at the labelled-end could facilitate the stabilization of a dimer-mediated synapsis. Alternatively, a single monomer forming a binary or pre-ternary complex at one end could be stabilized by the 5'-P present at the other end (Fig. 71A).

To further address this issue, we carried out DNAase I footprinting experiments with a 1 nt-gapped DNA substrate (Fig. 72A). In agreement with previous functional and structural data (Pitcher et al., 2007), DNase I footprinting analysis indicated that a 5'-P group is essential to stabilize the binding of Mt-PolDom to the gapped substrate. The length of the substrate protection caused by Mt-PolDom covers 3 bp to the downstream side, and 5 bp to the primer side. Thus, including the templating base, the footprinting covers a total of 9 nt of the template oligonucleotide. It is striking that exactly the same

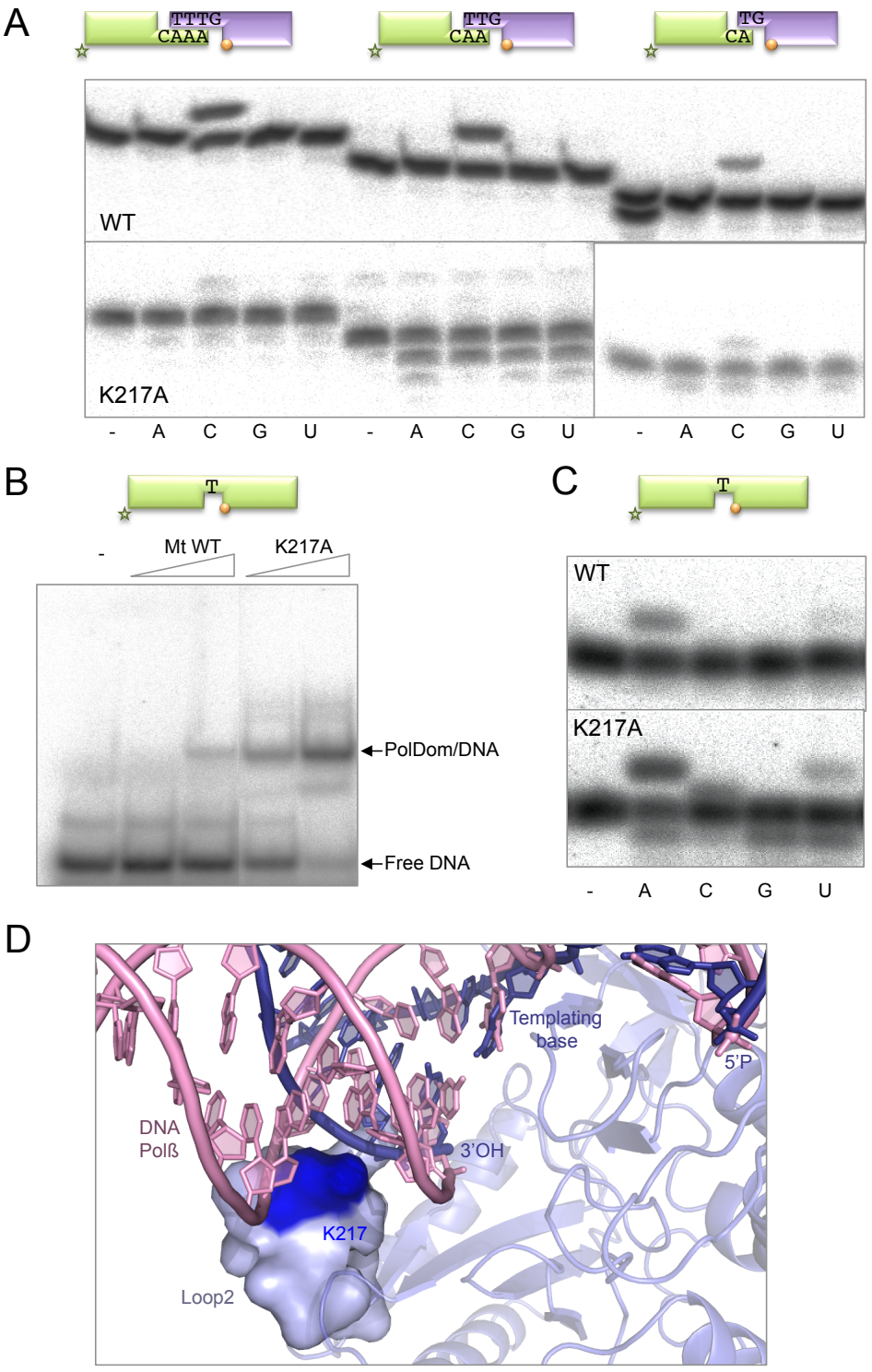



**Figure 73. Role of Loop2 in the stabilization of the primer.** A) NHEJ reactions were performed with 600 nM PolDom using various substrates formed with the oligonucleotides 1TTTG, 1TTG 1TG, 1AAAC, 1AAC or 1AC with 1D-NHEJ. The orange balls indicate the presence of a 5'-P group in the substrate. The star shows the position of the radioactive label. When indicated, each of the four NTPs (100 $\mu$M) were added in the presence of 1 mM $MnCl_2$. B) EMSA (left panel) was performed for the indicated proteins (200 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. After electrophoresis, gel was dried and the labeled fragments were detected by autoradiography. C) Gap-filling reactions (right panels) were peformed as described in Materials and Methods for the indicated proteins (25 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. When indicated, NTPs were added separately at 10 nM in the presence of 1 mM $MnCl_2$. D) Cartoon representation of the crystal structure of the PolDom complementary synapsis, in which the protein is shown in light blue, PolDom's own DNA substrate is shown in dark blue sticks, and the superimposed Polß DNA is shown in pink. Loop2 is shown in light blue surface, with the surface of $Lys^{217}$ highlighted in dark blue.

footprint is obtained with human Polß, a non-related monomeric enzyme involved in gap-filling reactions, suggesting that Mt-PolDom binds to gapped substrates as a monomer, and that the observed position of the DNA template and primer strands in a synapsis (see Fig. 73) is probably different when binding a gapped substrate, since only one monomer of PolDom has to protect this strand from DNAseI cut. Overimposition of the gapped-substrate crystallized with Polß on the new microhomology-mediated synapsis Mt-PolDom structure shows the possible new location of the upstream portion of the substrate, that would be now covered, and thus footprinted, by one PolDom monomer (Fig. 72B). This footprint size could be compatible with NHEJ reactions involving very short protrusions (Fig. 72C), that could be handled either by a single monomer, or by a dimeric arrangement as that previously

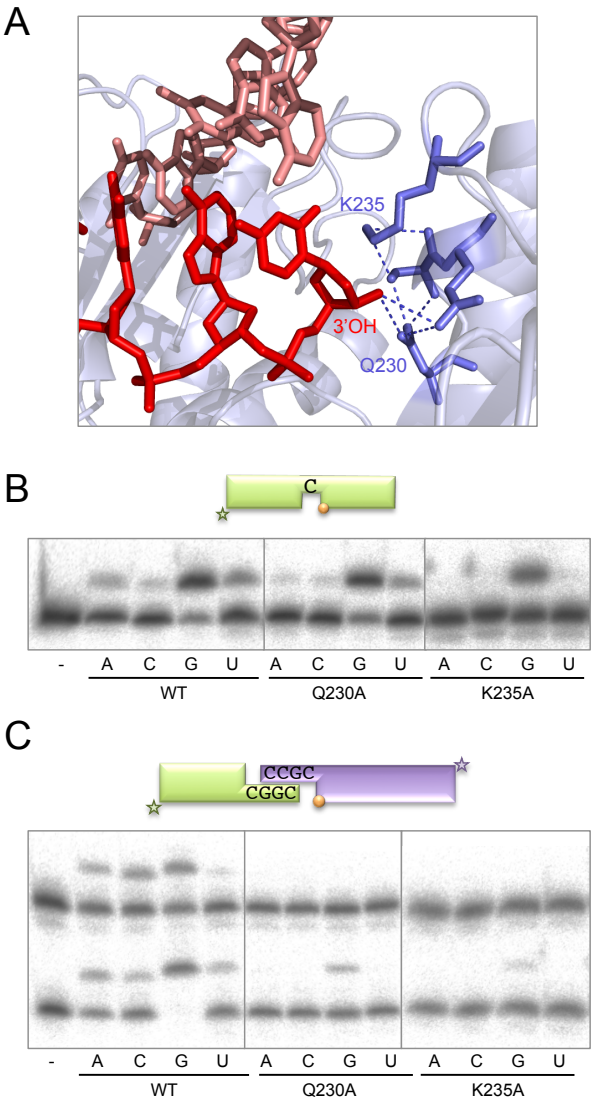



**Figure 74. Residues contacting the primer strand: specific role in NHEJ.**
A) Cartoon representation of the crystal structure of the PolDom complementary synapsis, in which the protein is shown in light blue, the DNA substrate is shown in red (primer) or pink (template) sticks, and the selected residues in blue sticks. B) Gap-filling reactions were peformed as described in Materials and Methods for the indicated proteins (25 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. When indicated, NTPs were added separately at 10 nM in the presence of 1 mM $MnCl_2$. C) NHEJ reactions were performed with 600 nM PolDom using a set of substrates formed with the oligonucleotides D3-C with D1 (green) and D4-C with D2 (mauve). The orange balls indicate the presence of a 5'-P group in the substrate. The star shows the position of the radioactive label. When indicated, each of the four NTPs (100 $\mu$M) were added in the presence of 1 mM $MnCl_2$.

described (Brissett et al., 2007).

## 12. Stabilization of the primer during synapsis: role of Loop 2 and other specific interactions with the primer terminus.

As we have recently proposed (Brissett et al., 2011), once a pre-ternary complex has been formed, including the template/downstream DNA substrate and the incoming nucleotide, the missing player that has to arrive and be stabilized by PolDom is the 3'-protruding DNA substrate acting as primer. A close examination of the structural elements of PolDom, in particular the Loop 2 region (residues 213-224), showed that Loop 2 exists as a $3_{10}$ helix in all of the determined Mt-PolDom structures except in the pre-ternary complex, in which the helix unravels and adopts a random coil conformation. This significant conformational change results in Cα position shifts of up to ~6Å, inducing a significant repositioning of two conserved residues, $Lys^{217}$ and $Arg^{220}$. The role of $Arg^{220}$ has been already discussed as a main player in the catalytic mechanism (Brissett et al., 2011), but the importance of the highly conserved $Lys^{217}$ remained uncertain. This positively charged residue contacts the 3'-OH of G13 (template strand) in the PolDom:DNA complex featuring an imperfect synapsis of two DNA ends (Brissett et al., 2007), and is also implicated in maintaining the position of the incoming primer in the fully complementary synapsis presented here. To establish the role played by $Lys^{217}$ in NHEJ-related DNA recognition processes, we mutated it to alanine and assayed its activity on NHEJ substrates. K217A mutation significantly affected the ability of the enzyme to promote synapsis (Fig. 73A). On the other hand, elimination of $Lys^{217}$ increased the ability of the enzyme to bind a gapped-DNA substrate (Fig. 73B), and concomitantly, its activity on this kind of primer-containing substrates was higher (Fig. 73C). A more in depth analysis of the structures available showed that the role of this residue during NHEJ can be explained through the stabilizing interactions stablished between $Lys^{217}$ and the incoming primer strand (Fig. 73D). On the other hand, no structure of Mt-PolDom binary or ternary complex bound to a gapped DNA substrate is still available, but taking into account the predicted location of the primer strand in such a complex (Figs. 72B and 73D), a steric hindrance between this upstream part of the substrate and $Lys^{217}$ in Loop 2 would surely occur. This could lead to an impaired binding to gapped

substrates, unbeneficial for ulterior catalysis. In the mutant, this negative clash/interaction is prevented, resulting in a strong positive effect in both DNA binding and activity of the enzyme. This evidence suggests a very precise role of Lys$^{217}$ in facilitating the more difficult connection of two separated 3'-ends (*via* NHEJ), while its role on more straightforward substrates, i.e. a DNA gap, is not only futile, but even detrimental for the achievement of the reaction. On that regard, it is tempting to speculate that some portion of Loop2, containing Lys$^{217}$, can be involved in a conformational change to maximize the capacity of the enzyme to bind and repair DNA gaps at a later stage of the NHEJ process.

Analyzing the structural similarities between PolX family members, we identified positively charged amino acids around the active site, which are potential stabilizing ligands for the 3'-nucleotide of the incoming primer strand. The 3'-OH itself is an axial ligand to the second catalytic metal and the phosphate group of the 3'-base is normally ligated by a nitrogen containing side-chain (Arg$^{254}$ in Polß, Arg$^{448}$ and Trp$^{342}$ in Polλ, or Arg$^{418}$ and His$^{329}$ in Pol*μ*). Analyzing the available structures of Mt-PolDom in complex with DNA, we found that several residues are forming a network that keeps the 3'-OH terminus of the primer in place for catalysis. The residues that are directly contacting this primer terminus are Gln$^{230}$ and Asp$^{227}$, while Lys$^{235}$ and Ser$^{229}$ are secondary players that form part of this interconnected web (Fig. 74A). In order to stablish the role of this network in the different activities of Mt-PolDom, mutants Q230A and K235A were obtained, purified and tested for polymerization on different substrates. Quite unexpectedly, these mutants maintained a high level of activity on a gapped substrate, as shown in figure 74B. The futility of this primer stabilizing interactions on this kind of substrates is probably due to the already fixed position of the primer through interactions with the template strand. To verify this possibility, we tested the mutants in NHEJ assays in which the primer needs to be stabilized by the polymerase since the template is a discontinuous strand. As expected, and even on substrates displaying a high level of complementarity, both Q230A and K235A showed almost undetectable levels of nucleotide incorporation when compared to the wild type PolDom (Fig. 74C).

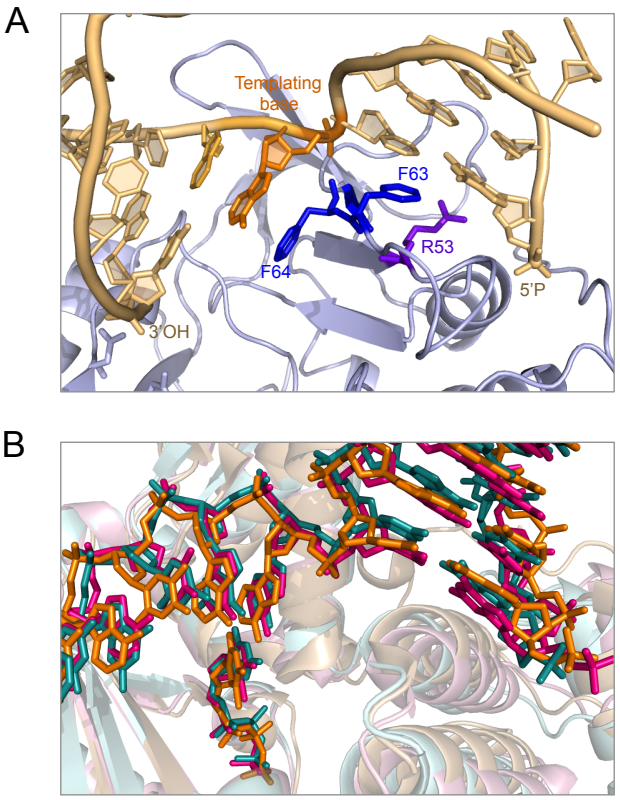



**Figure 75. Residues implicated in the templating base selectivity and maintaining the kink in the template strand.** A) Cartoon representation of the PolDom complementary synapsis with the protein shown in light blue, the DNA substrate shown in yellow with the templating base in orange, and the residues selected for mutagenesis in dark blue and purple sticks. B) Superimposition of the ternary complexes of the human Polß (pink), Polλ (teal) and Pol*μ* (orange) with the DNA substrates in sticks, showing the kink in the template strand.

All this data indicates that although Mt-PolDom can polymerize on gapped substrates, occurring during the second step of a NHEJ reaction and also in the course of other DNA repair pathways, the main function for which these polymerase is specially designed is the first stage of end-joining. The ability to bind strongly the 5'-P of the downstream strand together with the 3' primer terminus allow Mt-PolDom to perform its most unique function during repair of DSBs.

## 13. Adjusting the templating base for optimal catalysis.

The N-terminus of Mt-PolDom (residues 1-26) appears to be specific to the bacterial LigD-associated polymerase/primase family and is notably absent from the archaeal and eukaryotic primases (Brissett et al., 2011) suggesting that this region may be important for the NHEJ functions of

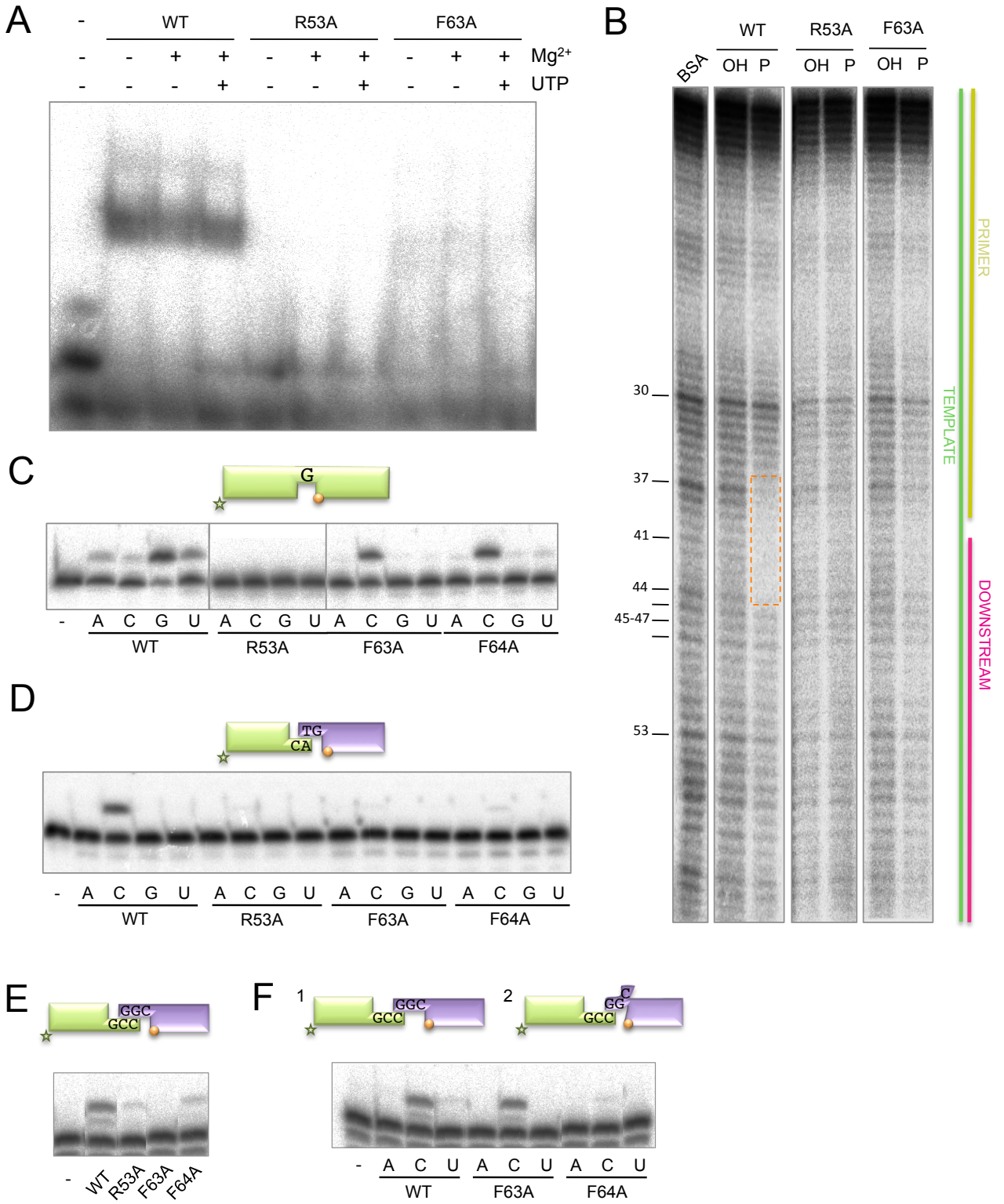



**Figure 76. Residues contacting the template strand: implications for PolDom-mediated NHEJ reactions.**
A) EMSA (left panel) was performed for the indicated proteins (200 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. When indicated, 1 mM $MnCl_2$ and/or 100 $\mu$M UTP were added. After electrophoresis, gel was dried and the labeled fragments were detected by autoradiography. B) Footprinting assay of wild-type or mutant PolDom (5 $\mu$g) was conducted as described in Materials and Methods. 10 $\mu$g of BSA were added to the control lane. C) Gap-filling reactions were peformed as described in Materials and Methods for the indicated proteins (25 nM) using a gapped substrate containing the oligonucleotides SP1C, T13C and DG1-P. When indicated, NTPs were added separately at 10 nM in the presence of 1 mM $MnCl_2$. D) NHEJ reactions were performed with 600 nM of the indicated proteins using a set of substrates formed with the oligonucleotides 1TG or 1AC with 1D-NHEJ. The orange ball indicates the presence of a 5'-P group in the substrate. The star shows the position of the radioactive label. When indicated, each of the four NTPs (100 $\mu$M) were added in the presence of 1 mM $MnCl_2$. E&F) NHEJ reactions were performed with 600 nM of the indicated proteins using a set of substrates formed by hybridizing the oligonucleotides CCG with 1D-NHEJ and GGC with 2D-NHEJ. The orange balls indicate the presence of a 5'-P group in the substrate. The star shows the position of the radioactive label. In E), only GTP (100 $\mu$M) was added in the presence of 1 mM $MnCl_2$, while in F) the other three were added (100 $\mu$M).

these enzymes. Many of the most important DNA binding residues (Asn$^{13}$, Lys$^{16}$ and Lys$^{26}$) reside here, where they form a positively charged phosphate-specific recognition pocket on the surface of the enzyme. A single mutation of Lys$^{16}$ is enough to largely abolish interaction with a 5´ P-containing DNA end (Brissett et al., 2007). Other conserved residues, including Arg$^{53}$, Phe$^{63}$, and Phe$^{64}$, also make direct contacts with each DNA end (Brissett et al., 2007), but their specific relevance in end-recognition and synapsis remained to be established, as they could have a more specific role in the precise alignment of the two ends after productive synapsis. Together, these residues, in addition to Glu$^{65}$ and Pro$^{55}$, form a molecular "wedge" that distorts the end of the DNA. These intimate contacts appear to play an important role in maintaining the kinking of the template strand by ~90° at the ds-ss junction (Fig. 75A). Amazingly, this orientation is comparable to that observed in Polß, Polλ and Pol*μ*, when complexed with gapped DNA (Fig. 75B).

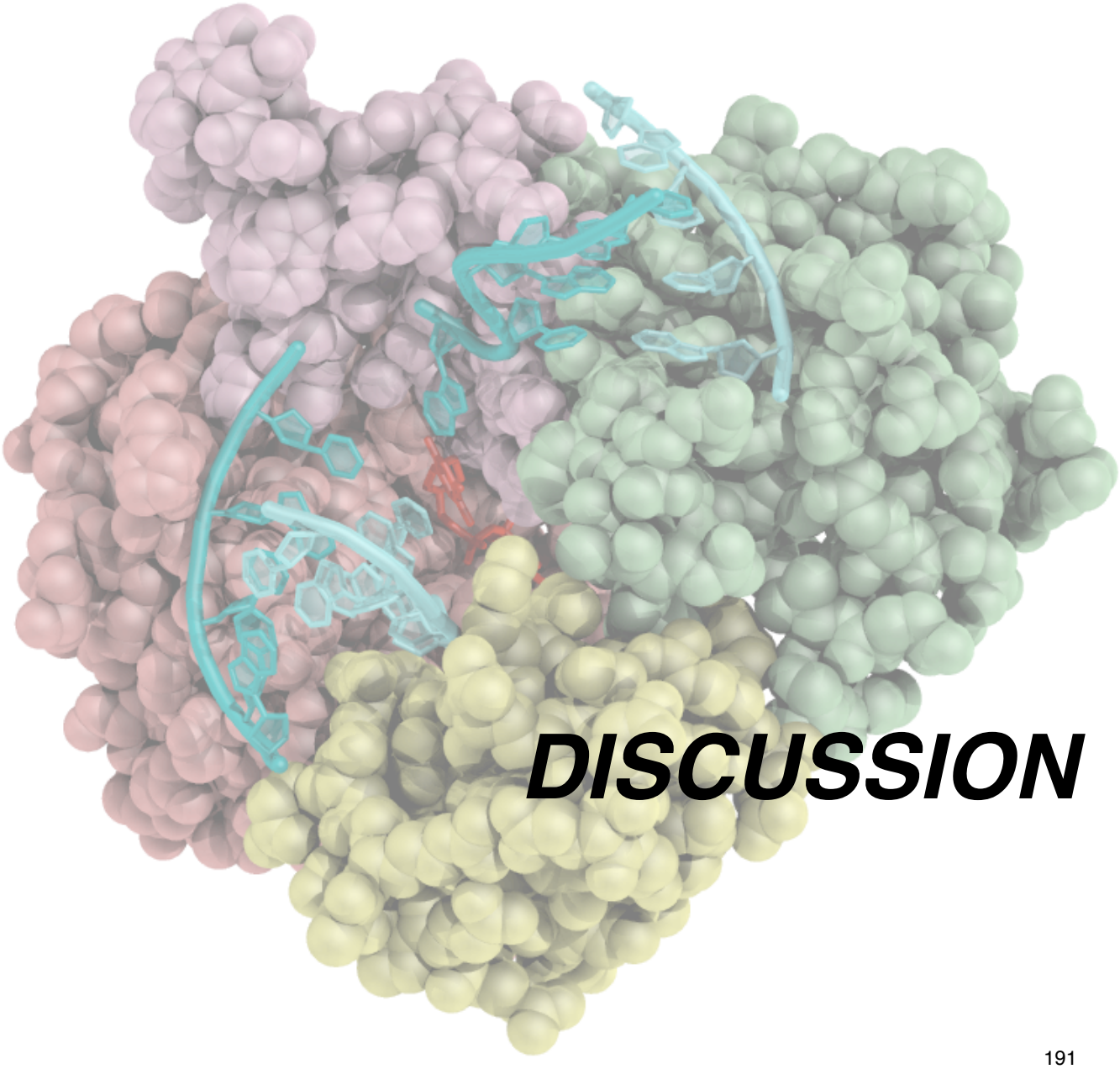
To establish the roles of residues Arg$^{53}$ and Phe$^{63}$, these amino acids were mutated to alanine and assayed to compare their ability to form a stable complex on gapped substrates. Both mutants R53A and F63A were very inefficient at binding to a gapped DNA substrate as assessed by EMSA (Fig. 76A), even in the presence of metal and/or nucleotide. The same results were obtained by footprinting analysis (Fig. 76B). In order to understand the relevance of these residues when binding to a gapped substrate we have to take into account that Arg$^{53}$ is involved in various interactions with Phe$^{63}$, Glu$^{65}$ and the 5´-terminal nucleotide of the downstream strand. Phe$^{63}$ is also involved in interaction with the base of the nucleotide (template strand) contributing to the first downstream base pair (containing the phosphate that limits the 5´-side of the gap), and together with Phe$^{64}$, is likely crucial to maintain the kink in the T/D DNA substrate. So, it can be inferred that the network of interactions formed by Arg$^{53}$, Phe$^{63}$, Phe$^{64}$, etc., has the maximal importance to adjust a proper binding to the gapped DNA substrate. In agreement with the observed reduction in DNA binding, gap-filling activity of these mutants was either barely detectable (R53A) or sensibly reduced (F63A and F64A) when compared to wild type PolDom (Fig. 76C).

Presumably, this fine adjustment of the templating base should be even more important in the case of NHEJ substrates, where the lack of a continuous template strand increases the need for stabilizing and orienting protein contacts. We decided to test the mutants R53A, F63A and F64A for polymerization on NHEJ substrates. As expected, the activity of the mutants when compared to the wild type enzyme was largely reduced on substrates with only one base pair of complementarity (Fig. 76D).

When we tested the mutants on NHEJ substrates having a higher complementarity at their 3'-protrusion (2 dG:dC bp), leaving a single templating base at each side of the connection, the levels of GTP incorporation displayed were either null (F63A) or barely detectable (Arg$^{53}$ and Phe$^{64}$), compared to the wild type enzyme (Fig. 76E). Interestingly, an unexpected outcome of this experiment confirmed the capacity of the wild-type PolDom to use alternative templating bases during NHEJ, as previously shown in a 2nt-gap context (Pitcher et al., 2007). In this case, the most favorable connection (2 dG:dC bp) between the two DNA ends will configure a 1nt-gap at each side of the synapsis (see scheme in Fig. 76E). Thus, the labeled primer (green) is expected to be extended only with G. However, when insertion of each of the four NTPs was tested, an equally efficient extension with C also occurred (Fig. 76F), and no reaction with ATP and UTP was observed (not shown), indicating that an alternative templating base (dG) is being used. This templating dG can be available if a single dG:dC base pair is sufficient for the connection, thus configuring a gap of 2nt (Fig. 76F, scheme 1). Recent results demonstrated that the templating base closest to the 5'-P is the preferred one to select an incoming nucleotide, even in the absence of a primer strand (Brissett et al., 2011). For CTP selection to occur, the preceding templating base (dG) must substitute the favorite dC, but keeping contact with the 5'-P. A plausible explanation is that dC can be "scrunched", awaiting to be used in a next round of nucleotide incorporation (see Fig. 76F, scheme 2), a situation already described for human Polλ in a complex with 2nt-gapped DNA (Garcia-Diaz et al., 2009). This ability implies the existence of specific interactions with the scrunched base(s), thus allowing both polymerases "to count", consecutively reading several templating bases in a gap/NHEJ intermediate. In agreement with this model, it is

likely that residues Phe$^{63}$ and Phe$^{64}$ can have a crucial influence in the decision to select the templating base also in these situations. Strikingly, mutant F63A, that is completely unable to use the favorite templating base (dC), could insert C as efficiently as the wild-type (Fig. 76F), being forced to select the "scrunching" option (scheme 2). Mutant F64A, conversely, was only able to catalyze insertion of G in front of the last templating base (dC; Fig. 76F), therefore not allowing the "scrunching" option (scheme in fig. 76E).

Therefore, we have shown that alternative bases at variable distances from the 5'-P can be selected as valid templates during NHEJ, and that this decision depends on the flexibility of the interaction guided by the phenylalanine pair (Phe$^{63}$-Phe$^{64}$). In general, choosing the "scrunching" option will minimize the connection needed and the loss of sequences flanking the break.



# DISCUSSION

## 1. General architecture of X family polymerases.

### 1.1 BRCT domain.

The members of the X family of polymerases are recruited to form a complex with the NHEJ core factors XRCC4/Ligase IV and Ku at the DNA break (Lee et al., 2004; Nick McElhinny et al., 2005; Tseng and Tomkinson, 2002). Recent evidence has shown that BRCT domains can be specifically involved in the interaction with phospho-serine or phospho-treonine containing motifs (Manke et al., 2003; Yu et al., 2003), an ability that may be involved in granting access of regulated proteins to the break, even though no evidence has shown to date a phosphorylation-dependent, BRCT-mediated, interaction of NHEJ factors.

Interestingly, sequence comparisons show that the BRCT of Pol*μ* is most similar to TdT, with 39% sequence identity that includes the residues important for complex formation (DeRose et al., 2007). The hight level of sequence conservation is translated to the structural level in the BRCT domains of Pol*μ* (PDB ID: 2DUN) and TdT (PDB ID: 2COE), that in turn exhibit an a/ß motif that is similar to the BRCT found in XRCC1 (PDB ID: 1CDZ), a BER repair protein. The main differences include a shorter a-helix 2 in the TdT BRCT domain, as well as the positioning of the loop connecting a-helix 2 and ß-strand 4. The electrostatic surfaces of Pol*μ* and TdT BRCT domains are also very similar, containing both a positively charged ridge on one face of the protein, and large negatively charged regions on the opposite faces. In the Pol*μ* BRCT, the positive ridge is formed by $Arg^{44}$, $Arg^{52}$, $Arg^{85}$ and $Arg^{86}$. This positive patch has been proposed to be involved in the interaction with a phospho-modified protein (DeRose et al., 2007), but we have demonstrated that its most probable function is the interaction with the downstream part of the DNA substrate (this PhD Thesis). Point mutations in several residues of the positive ridge are being made and tested, but the complete lack of the domain already resulted in a diminished interaction with and activity on NHEJ substrates. By using the "YACQR" motif to correctly orient and overimpose the crystals of the BRCT domain and the Pol*μ* core,

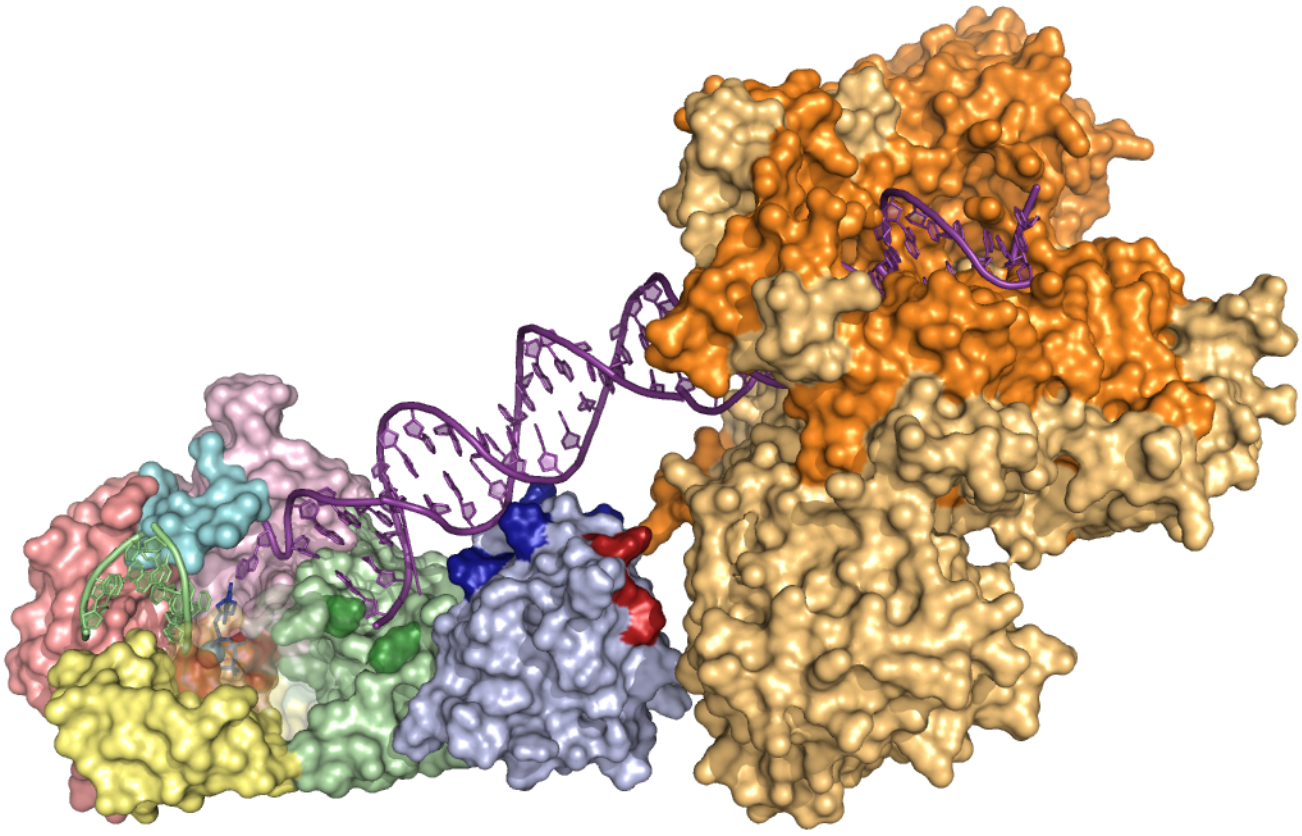


**Figure 77. Model of the interaction of Pol*μ* with the Ku heterodimer and the DNA substrate through the BRCT domain.**

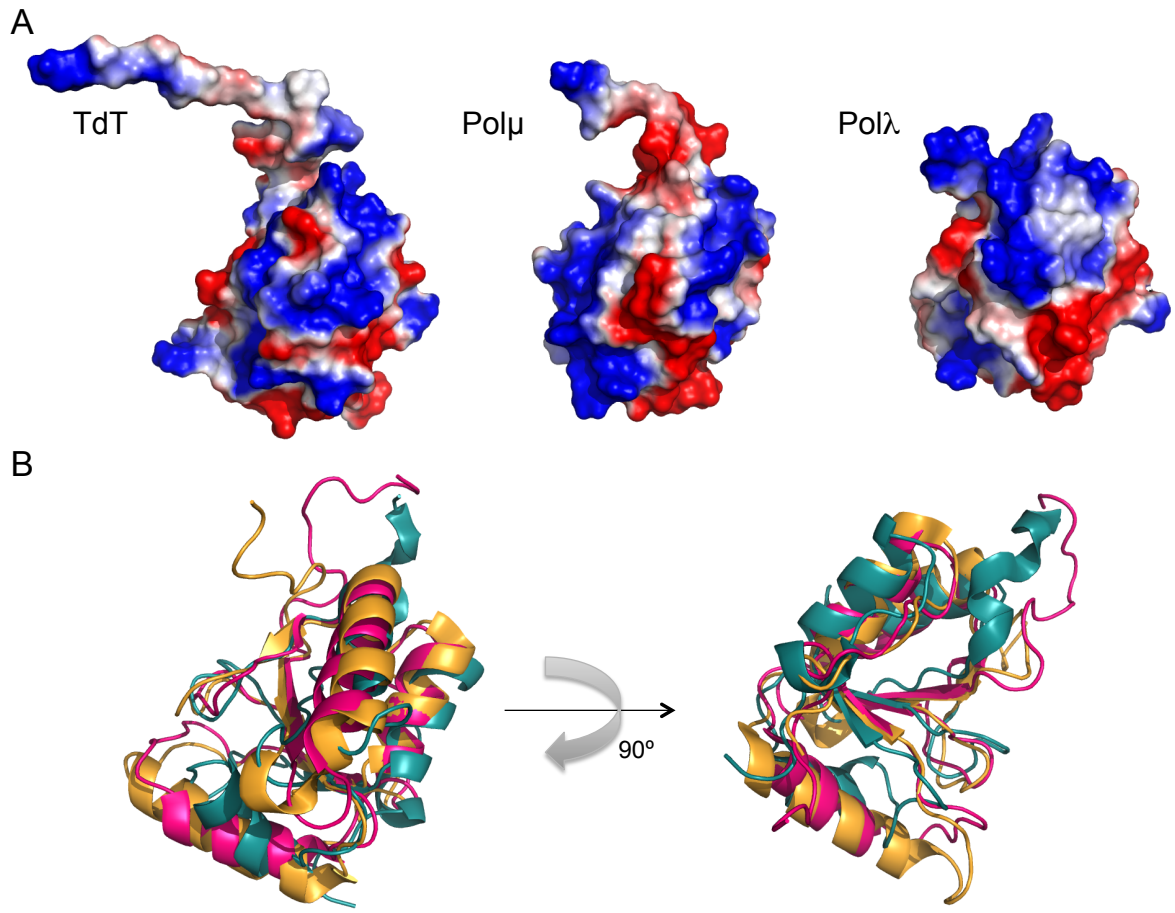



**Figure 78. BRCT domains of the human X family members.**
A) Electrostatic surface of the BRCT domains of TdT (2COE), Pol*μ* (2DUN) and Polλ (2JW5). B) Superimposition of the BRCT domains of the X family members: TdT (dark pink), Pol*μ* (orange) and Polλ (teal), shown in cartoon.

we found out that one of the positive patches in the BRCT domain (Fig. 37, Patch I) perfectly accommodates the downstream part of the DNA substrate (Fig. 77; colored in dark blue). We then modeled the interaction of the BRCT domain of Pol*μ* with the Ku70/Ku80 heterodimer by orienting the DNA substrate. Strikingly, the side of the BRCT domain facing the Ku heterodimer in the model was exactly the one containing the residues reported to be involved in this interaction (Fig. 77; colored in red). According to this model, the portion of the DNA substrate which would be contacted by the BRCT domain flawlessly correlates with the length of the BRCT-specific protection (6 bp) observed in our footprinting assays.

This DNA binding function of Pol*μ* BRCT, independent of the core NHEJ factors, may enable this polymerase for a role in the alternative NHEJ pathway, which occurs independently of Ku or Ligase IV. Pol*μ* might bind the DNA break based on its own specificity for the 5'-P and through its BRCT domain, and *via* its terminal transferase be in charge of the additions that create the so called polymerase-generated microhomology. In agreement with this proposed function, recent observations indicate that Pol*μ* BRCT is atypical in the sense of not being involved in dimerization or multimerization. In fact, comparison of the structure of Pol*μ* BRCT with other BRCT domains which effectively dimerize, shows important differences, specially regarding R2 helix (Mueller et al., 2008).

The sequence conservation among BRCT domains from family X polymerases is very low, with only 10 residues conserved and five of them ($His^{82}$, $Val^{84}$, $Leu^{109}$, $Trp^{114}$, and $Leu^{115}$, Polλ numbers) involved in the architecture of the domain. The other five ($Gly^{54}$, $Arg^{57}$, $Gly^{69}$, $Thr^{81}$, and $Val^{125}$) are exposed to the solvent in the surface of the protein. One of them, $Arg^{43}$ in Pol*μ* ($Arg^{57}$ in Polλ) has been shown to be implicated in interactions with other components of the NHEJ complex (DeRose et al., 2007).

This lack of sequence homology is reflected in structural variations of the family X polymerases BRCT domains (Fig. 78), which in turn influence the interactions stablished with other NHEJ factors, including an improved/preferential access of the

polymerase to the DNA break. Deletion of the BRCT domain in the NHEJ-related polymerases (Fan and Wu, 2004; Nick McElhinny et al., 2005; Tseng and Tomkinson, 2002), or point-mutagenesis of key-residues (DeRose et al., 2007; Mueller et al., 2008), block the formation of complexes between the polymerase, Ku and XRCC4/LigaseIV at DNA ends.
The ability of X family polymerases to act during classical NHEJ thus relies on their interactions with other NHEJ factors through their BRCT domains, but PolXs have intrinsic capacities of gap-recognition and binding involving simultaneous recognition of both sides of the gap. As shown for Polß, the polymerase can bind both the template/primer part of the gap and also the template/downstream part, being the latter the part that is most strongly bound. In the Polß co-crystal with a DNA gap this dual binding is clearly observable, stablishing contacts with the DNA backbone through a positively charged platform onto which the DNA is leaning (Fig. 78). Such a dual DNA binding is more crucial for Polλ and Pol*μ*, polymerases not as specialized as Polß in always confronting continuous substrates (i.e. gaps), but also in charge of bridging two separate DNA ends. The ability to independently bind and orient two DNA ends is thus closely related to their function during NHEJ, and is still found in a more recently evolved enzyme, Polß, as an appropriate solution for gap-filling. This tight binding to both sides of the templating base forces the formation of a sharp bend of 90º in the template strand, that has been proposed to increase nucleotide selectivity and sensitivity to mismatches, and in general be a mechanistic feature used by X polymerases in order to improve fidelity (Pelletier et al., 1996).

**1.2 Serine/proline domain.**

In the case of Polλ, the catalytic core domain and the BRCT domain are connected by a segment of about 100 amino acid residues that is rich in serine, threonine and proline residues (Ser-Pro domain). This flexible linker is much shorter in TdT and Pol*μ*, although also contains a high level of serines and prolines, and is absent in Polß. Structural predictions suggest that the flexibility and accessibility of this random-coiled region in Polλ would be adequate to mediate or coordinate protein:protein contacts or conformational changes required for its function. Moreover, this region is predicted to contain multiple targets for phosphorylation and glycosylation (Garcia-Diaz et al., 2000). Interestingly, the similarity between human and murine Polλ is the lowest (55% amino acid identity) in the Ser-Pro domain (residues 126-238 of murine Polλ and residues 127-240 of human Polλ). Despite an overall lesser similarity, the conservation in this domain appears to be more restricted to serine, threonine, and proline residues, thus emphasizing the putative role of this region as a target for phosphorylation (Garcia-Diaz et al., 2002). In agreement with this, three serines in the Ser-Pro region of Polλ (amino-acids 167, 177 and 230) form part of consensus sites for CDK phosphorylation: kinase assays performed with deletion mutants lacking the Ser-Pro domain show that this region is the target for *in vitro* phosphorylation by Cdk2/CycA (Frouin et al., 2005). This regulatory function of the Ser-Pro domain could be conserved in Pol*μ*, as the region connecting the BRCT and the 8 kDa domain in Pol*μ* sequence, although much smaller in size, it contains some consensus sites for CDK phosphorylation.
Another line of evidence implicates the Polλ Ser-Pro domain in the regulation of the activity of the polymerase domain: firstly, a deletion mutant of the first 244 amino-acids (BRCT+Ser-Pro domains) has been shown to display a much higher polymerase activity and stronger strand-displacement that the wild-type Polλ (Fan and Wu, 2004); secondly, pre-steady-state kinetics comparing full-length Polλ with a truncated form that again lacks the first 244 amino-acids indicate that this Ser-Pro domain confers an increase in fidelity by lowering the incorporation rate constants of incorrect nucleotides, thus allowing Polλ to reach the fidelity levels of Polß (Fiala et al., 2006).
On the other hand, all the biochemical studies leading to the conclusion that the Ser-Pro domain is implicated in the regulation of the catalytic activity of the polymerase have been performed on deletion mutants which not only lacked the Ser-Pro (and BRCT) domain, but also a previously unnoticed region of the protein which has been studied in depth in Pol*μ* in this PhD thesis: the “closing” motif, conserved in Pol*μ*, TdT, and also Polλ. This motif, consisting of only five amino-acids (141-YACQR-145) in Pol*μ*, 239-WxCxQ-243 in Polλ), is implicated in maintaining the closed conformation of the catalytic core of these polymerases, as shown in TdT and Pol*μ* crystal structures, a function that is essential for Pol*μ*-driven NHEJ (this PhD thesis).

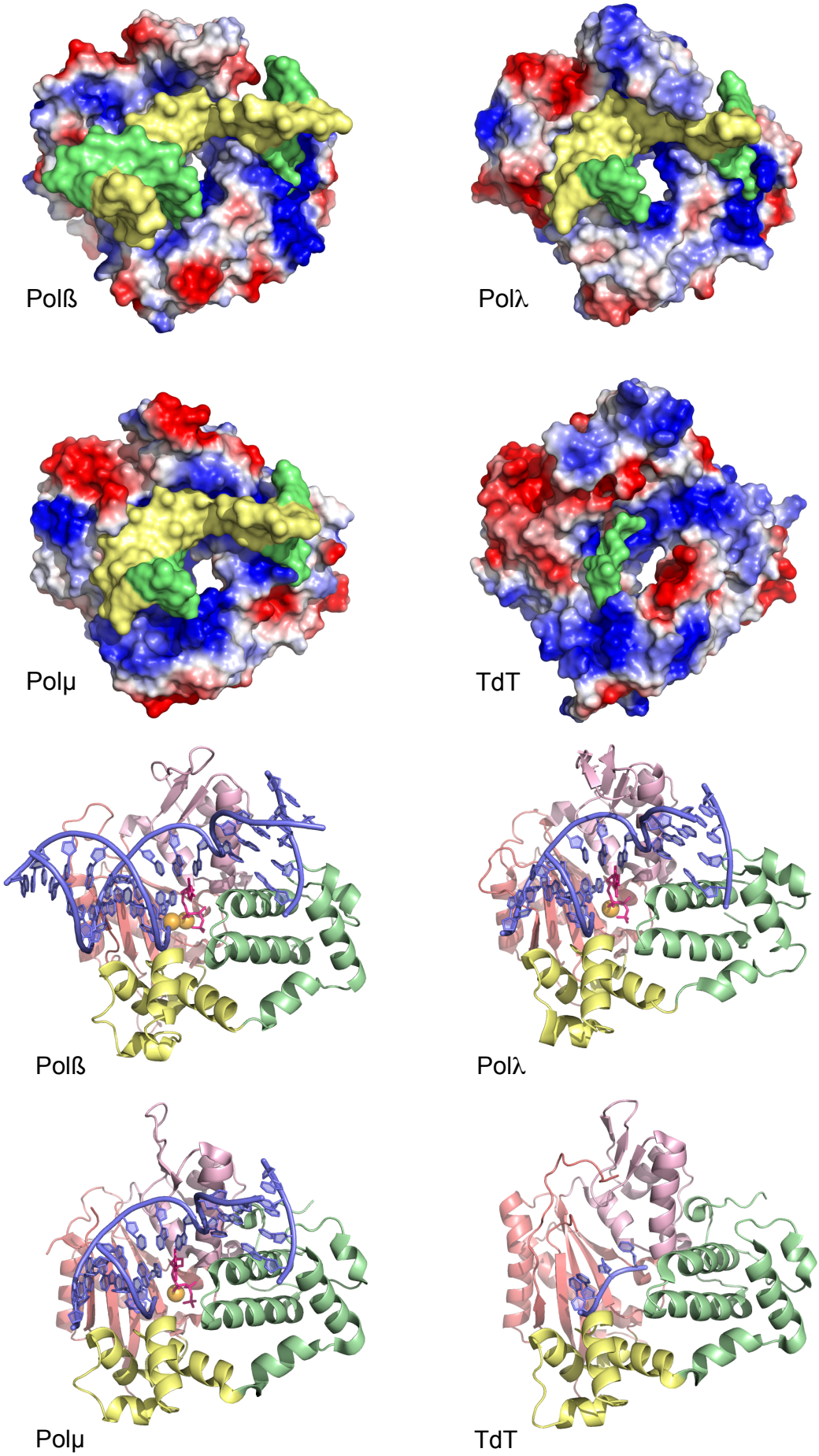



**Figure 79. Polymerase domains of the human X family members.**
Polymerase domains of Polß (1BPZ) TdT (1KDH), Pol*μ* (2IHM) and Polλ (1XSN), shown in electrostatic surface with the DNA substrates in green and yellow (top part) or in cartoon with the DNA substrates in blue sticks and the domains colored as follows: 8 kDa in green, fingers in yellow, palm in red, thumb in pink.

Unfortunately, in the case of all the structures solved for Polλ this small region is missing. However, this fact could well explain the 5-fold difference in $K_m$ observed between the full-length Polλ and the crystallized form of the enzyme (Garcia-Diaz et al., 2004) and also the conformation adopted by the 8 kDa domain in this Polλ structure, which is intermediate between the open and closed conformation of Polß (PDB IDS: 1BPZ, 1BPY).
This lack of the "closing" motif could also explain the results obtained by Fan and Wu, 2004; Fiala et al., 2006, with their truncated mutants, thus not allowing to delineate a specific role of the Ser-Pro domain from their data.

### 1.3 Polymerase domain.

#### *1.3.1 8 kDa domain.*

The unique structural feature that allows polymerases from X family to bind gapped and NHEJ substrates is the 8 kDa domain (Fig. 80A), located either at the N-terminus (Polß from higher eukaryotes, bacteria and archaea), or at the N-terminal portion located just after the flexible linker that contains the Ser-Pro domain (Polλ, Pol*μ*, TdT and yeast Pol4). This 8 kDa domain is involved in contacting several parts of the DNA substrate through different motifs (Moon et al., 2007), but in some of the members of the X family bears a dRP-lyase activity, highly related to the BER pathway (Garcia-Diaz et al., 2001; Prasad et al., 1998).
With the resolution of the first crystal structures of rat and human Polß, the 8 kDa domain was found to be highly mobile (Fig. 80B), not freely, displaying a small number of stable positions: 1) in the absence of DNA and incoming nucleotide, the 8 kDa domain is located far away from the thumb subdomain, and the polymerase is in an open conformation; 2) in the presence of a DNA gap, the 8 kDa domain moves and comes closer to the thumb through binding of

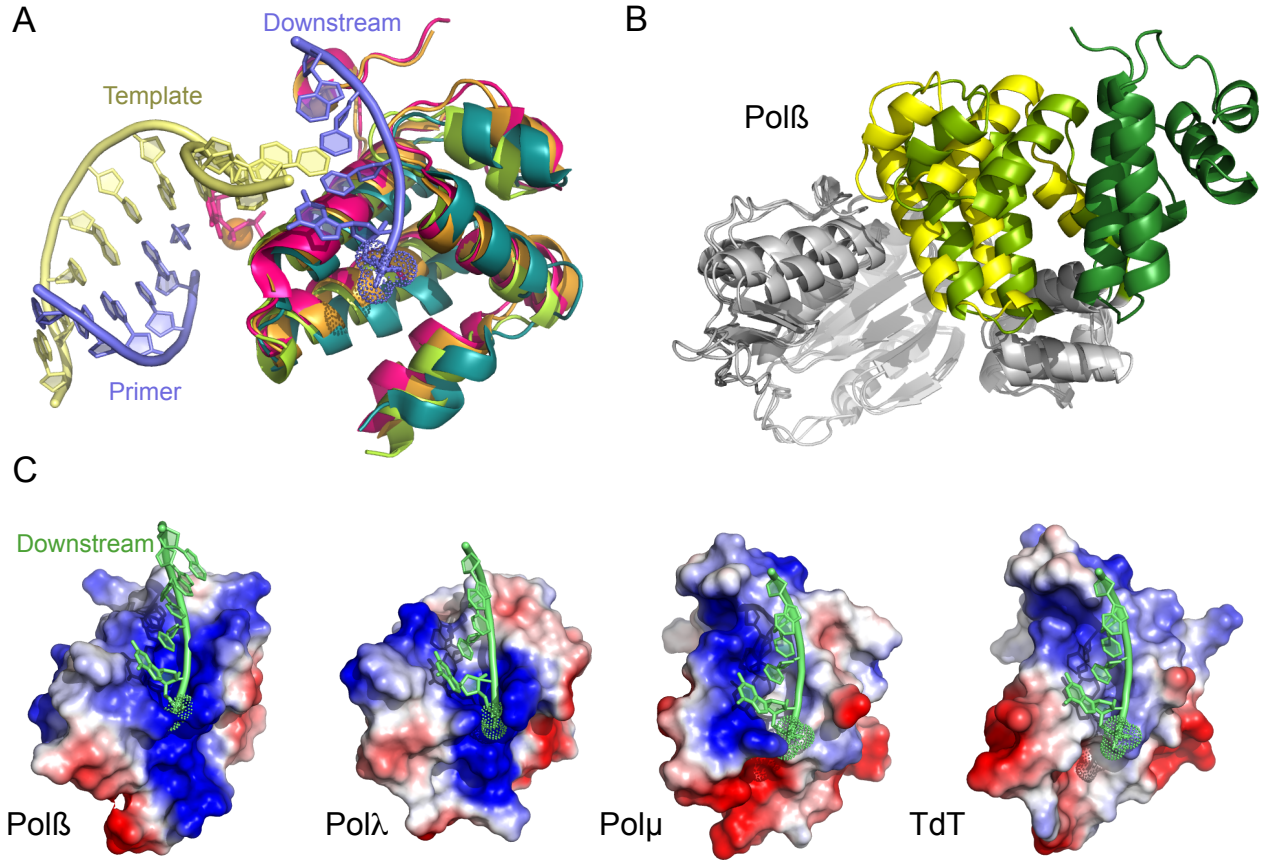



**Figure 80. 8 kDa domain of the human X family members.**
A) Superimposition of the 8 kDa domains of the X family members: Polß (dark pink), Pol*μ* (orange) and Polλ (teal), shown in cartoon. B) Superimposition of the structures of Polß: apoenzyme (1BPX, dark green), binary (1BPY), light green and ternary (1BPZ, yellow) complexes. C) Electrostatic surface of the 8 kDa domain of Polß (1BPZ), TdT (1KDH), Pol*μ* (2IHM) and Polλ (1XSN), with the downstream strand shown in green. Pol*μ* DNA was superimposed on the TdT structure.

the 5'-phosphate group of the downstream strand; 3) after arrival of the nucleotide, there is a further movement of the 8 kDa domain, and Polß finally adopts the closed conformation. The model proposed by Pelletier et al., 1996, explains the formation of the 90º bend in the DNA substrate in two steps: first, binding of the 8 kDa domain to the downstream part of the gap stabilizes the initial positioning of the enzyme; secondly, upon folding of the 31 KDa polymerase domain and binding of the primer part of the substrate, the bend of the DNA duplex is created. This bending causes the downstream part to rotate out, exposing the 3' end of the primer.

This two-step model is confirmed by the observations derived form the solved Polλ structures, the most indicative in this matter being the co-crystal with a 2 nt gap (Garcia-Diaz et al., 2004, PDB ID: 1RZT). In this case, the 5'-P is located in its correct position and bound by the 8 kDa domain, but the place of the templating base is occupied by the second template nucleotide of the gap, i.e. the one adjacent to the downstream duplex. This causes the 3'-OH of the primer to be displaced to the -1 position, relative to the catalytic position, adjacent to the NTP binding site, observed in the 1 nt gap co-crystal (ref, PDB ID: 1XSN). Therefore, the location of the polymerase domain in a gap (1-nt or longer) is dictated by the binding of the 8 kDa domain to the 5'-P, and not by interactions with the primer terminus.

This conclusion has implications of great interest for the binding of the polymerase to NHEJ substrates, mainly as 8 kDa-mediated binding would occur irrespective of the conformation of the 3' end. The polymerase in charge for this has to be able to take advantage of microhomologies for aligning the 3 ´ends, and the 8 kDa domain provides an anchoring point for this complicated task.

*1.3.1.1 Lyase activity.*

The 8 kDa domain of Polß and Polλ harbor an intrinsic dRP lyase activity that is required during single-nucleotide BER to remove the residual 5'-deoxyribose-phospate moiety left by the AP-endonuclease after elimination of the nitrogenous base. This reaction proceeds through a ß-elimination mechanism *via* a Schiff base intermediate, and has been shown to be the rate-limiting step in the elimination of several DNA lesions *in vivo* (Garcia-Diaz et al., 2001; Sobol et al., 2000). The studies on the structural aspects of dRP-lyase chemistry (Beard et al., 2006; Garcia-Diaz et al., 2005; Prasad et al., 2005) have led to the conclusion that the amino acids serving as catalytic nucleophiles are $Lys^{72}$ in Polß (Prasad et al., 1998) and $Lys^{312}$ in Polλ (Garcia-Diaz et al., 2001). This positively charged residue is not conserved in Pol*μ* ($Val^{212}$) or TdT ($Val^{224}$), and thus the dRP-lyase activity is not present in these enzymes.

*1.3.1.2 Phosphate pocket.*

As already noted, the main function of the 8 kDa domain is the binding of the 5'-P group of the downstream strand of the DNA substrate. In fact, polymerization rates by template-instructed polymerases of the X family are greatly enhanced when the substrate contains this 5'-P group. In the case of Polß and Polλ, the processivity is also improved on long gaps (5 nt; Garcia-Diaz et al., 2002; Singhal and Wilson, 1993). In the ternary structures of Polß (PDB ID: 1BPY), Polλ (PDB ID: 1XSN) and Pol*μ* (PDB ID: 2IHM) this 5'-P moiety is located at a positively charged pocket where binding is mediated by several hydrogen bonding interactions with basic side chains within the pocket (Fig. 80C). However, in Pol*μ* there are fewer interactions than in Polß or Polλ, and the binding pocket is not as positively charged (Fig. 80C). There is no structure of TdT containing a downstream strand, but this enzyme still conserves the 8 kDa domain, that could be used to coordinate terminal addition of N-nucleotides with the joining of the two DNA ends generated during V(D)J recombination.

*1.3.1.3 HhH domain.*

The 8 kDa domain contains another structural motif implicated in DNA binding, the helix-hairpin-helix (HhH) motif. This motifs bind single- or double-stranded DNA in a sequence independent manner, with the aid of a coordinated metal cation (Doherty et al., 1996; Mullen and Wilson, 1997). In Polß, Polλ and Pol*μ* structures, this HhH interacts with the downstream part of the substrate, suggesting that its function is the stabilization of the bent DNA, thereby facilitating the positioning of the two DNA ends in a NHEJ reaction.

The structures of the 8 kDa HhH motifs from the X family enzymes are not exactly the same: in Polß and Polλ this motif is similar to those found in other

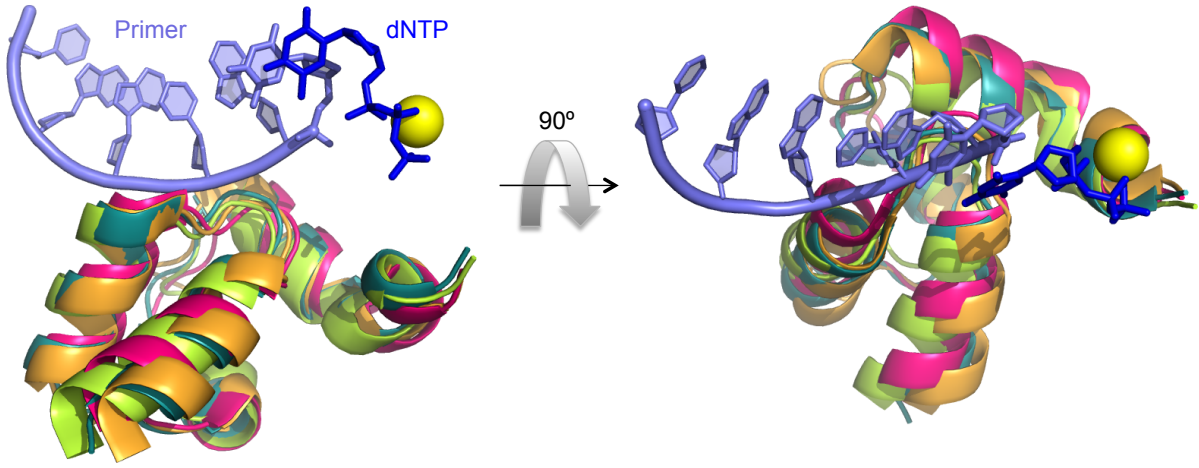



**Figure 81. Fingers domain of the human X family members.**
Superimposition of the fingers domain of the X family members: Polß (light green), TdT (dark pink), Pol*μ* (orange) and Polλ (teal), shown in cartoon. DNA substrate shown in light blue, incoming dNTP shown in dark blue, metal ion shown as a yellow sphere.

repair enzymes, with the GxG sequence of the hairpin and other protein residues being conserved. In Pol*μ* and TdT, on the other hand, one of the helices is distorted, probably a consequence of the lack of primary sequence conservation in the hairpin (CLG in TdT, HFG and YLG in mouse and human Pol*μ*, respectively).

### *1.3.2 Fingers subdomain.*

One of the main functions of the fingers domain in the X family polymerases is to provide a platform for the primer strand (Fig. 81): a second HhH in this region (α-helices F and G) interacts with the upstream duplex. This HhH motif utilizes a metal ion in order to act as a kind of processivity factor by increasing the affinity of the polymerase for the substrate and helping the translocation along the DNA backbone. Two different models were proposed by Pelletier and Sawaya, 1996, to explain the mechanism by which the metal ion could be used: either the metal remains tightly bound by the amino-acids side chains and “rolls” (breaking and creating bonds) over the DNA backbone, or the metal remains strongly bound to the DNA and it is the polymerase that “travels” along this “magnetic railing”, that is renewed with each translocation event.

In Pol*μ*, there are only two putative hydrogen-bonding interactions from the fingers subdomain to the template strand, one of which is mediated by a bridging water molecule. Conversely, the interactions between the HhH motif in the fingers domain and the upstream primer are more numerous and span a larger region than those found in Polλ. The phosphate backbone of the primer is accommodated in a pocket created by protein residues $Gly^{247}$–$Arg^{253}$ (including the hairpin region, GVG) and is positioned by both non-bonded and hydrogen-bonding interactions. These interactions may allow the protein to align the primer within the DNA-binding cleft, specially in the case of NHEJ substrates or in situations in which a special primer-binding strength is needed, such as the filling-in of longer gaps in the presence of a continuous or discontinuous template strand (this PhD thesis).

Accordingly with the functions proposed for the HhH motifs in X family polymerases, the only polymerase lacking the two described HhH motifs, ASFV pol X, retains a very limited ability to carry out processive synthesis, and only on very short (3 nucleotide) 5’-phosphate-gapped DNA molecules (Oliveros et al., 1997).

### *1.3.3 Palm subdomain.*

In view of the first crystal structure of rat Polß (Sawaya et al., 1994), the similarity among the different polymerases crystallized to that date (Klenow fragment, reverse transcriptase, RNA polymerase) and, specially, of the palm subdomains, was obvious. The basic structure of the palm includes two α-helices packed against one face of a β-sheet (Fig. 82A). Such a fold is common in many proteins and has been termed a two layered α-ß sandwhich (Sousa et al., 1993). The main difference

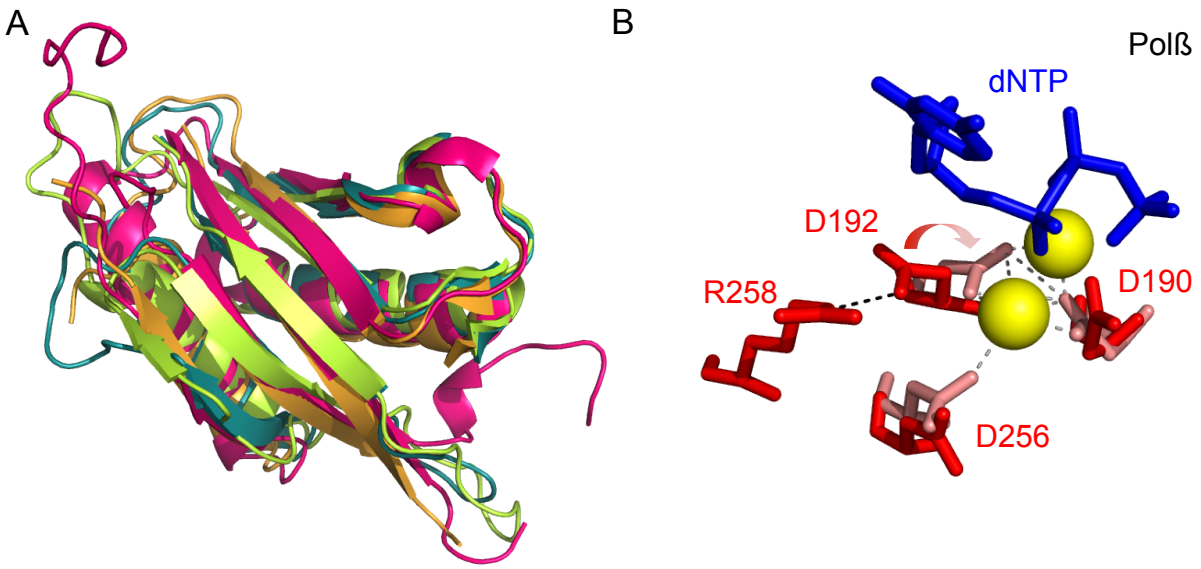



**Figure 82. Palm domain of the human X family members.**
A) Superimposition of the palm domain of the X family members: Polß (light green), TdT (dark pink), Pol*μ* (orange) and Polλ (teal), shown in cartoon. B) Superimposition of the three catalytic aspartates of Polß from the binary (red sticks) and from the ternary (pink sticks) complexes. Incoming dNTP shown in dark blue, metal ions shown as yellow spheres.

in Polß palm is that it contains an additional inserted parallel strand that supports one of the catalytic aspartates. The other two catalytic residues are also located in the palm. These residues, always located in the palm subdomain, form a carboxylate triad implicated in the arrangement of a pair of divalent metal ions in a precise geometry to catalyze the nucleotidyl transfer reaction. In the case of Polß, one of these carboxylate residues adopts two different conformations, depending on the presence or absence of the incoming nucleotide: in the open conformation it is hydrogen-bonded to an arginine in a conformation that prevents interaction with the metals. Upon binding of the correct nucleotide, this interaction is disrupted, allowing a change in the conformation of the aspartate, such that it can coordinate the active site metals (Fig. 82B). Another difference between X family enzymes and other polymerases is that their subdomains are sequentially continuous: one subdomain follows the other both in the tertiary and in the primary structure (Fig. 79). Although the topology of the palm subdomains of family X polymerases is not homologous to families A, B, Y, and RT polymerases, the catalytic participants (metals, dNTP, and DNA) can be structurally aligned (Steitz et al., 1994).

Regarding substrate binding, in the case of Pol*μ* there are extensive non-bonding and hydrogen-bonding interactions from the palm subdomain to the upstream region of the template strand, mostly in a pocket formed by residues $His^{366}$–$His^{369}$. Also, interactions between the palm subdomain and primer terminus are crucial for correctly positioning the nucleotide for catalysis. Most interactions between Pol*μ* and the upstream duplex are with the phosphate backbone. The exception is $Arg^{389}$, located in the minor groove. Also notable is that the position of the primer terminal nucleotide appears to be regulated by two putative hydrogen bonds, involving $His^{329}$ and $Arg^{420}$ (Moon et al., 2007).

Another specialty in the palm subdomain is Loop 1: a motif between beta strands 3 and 4 of the Pol*μ* and TdT palm that is more than 13 amino-acids longer than Polß or Polλ Loop 1. In a structure of TdT with a ssDNA primer (Delarue et al., 2002), Loop 1 is found in place of the template. This positioning of the TdT Loop 1 suggests an explanation for the TdT and Pol*μ* affinity for ssDNA and 3' overhang-containing substrates, and consequently for their terminal transferase activity. Therefore, Loop 1 is an important determinant of DNA substrate specificity during NHEJ of double-strand breaks in DNA, as will be discussed more in depth below.

### *1.3.4 Thumb subdomain*

The change in Polß from an "open" to a "closed" conformation upon nucleotide binding and back after catalysis is triggered by movements of the thumb

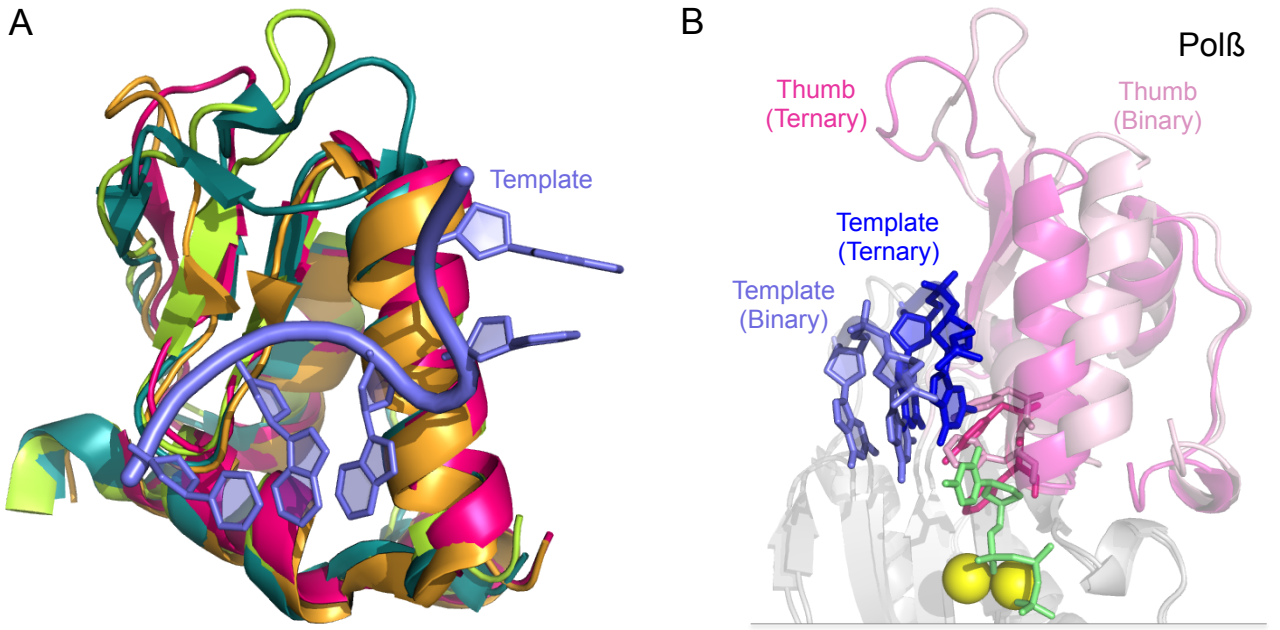



**Figure 83. Thumb domain of the human X family members.**
A) Superimposition of the thumb domain of the X family members: Polß (light green), TdT (dark pink), Pol*μ* (orange) and Polλ (teal), shown in cartoon. DNA substrate shown in light blue. B) Superimposition of the thumb domain of Polß from the binary (light pink) and from the ternary (dark pink) complexes, showing the DNA substrates from both structures (in light and dark blue respectively), the incoming dNTP shown in green and the metal ions shown as yellow spheres.

subdomain (Fig. 83B). Specifically, α-helix N is repositioned to directly interact with the nascent base pair (Beard and Wilson, 1998). This serves as a kind of testing of the geometric position of the incoming nucleotide and the primer terminus. When the thumb is open, the mononucleotide-binding motif rotates away from the active site, weakening interactions with the dNTP's phosphates. Thumb closure favors the active dNTP conformation by moving the mononucleotide-binding motif closer to the primer 3′-OH (Sawaya et al., 1997). The palm-thumb interface is lined with numerous hydrophobic groups on both sides of the hinge, and these hydrophobic interactions may favor both conformations of the thumb (Sawaya et al., 1997). This thumb subdomain motion is absent in Polλ. Instead, a shift of the DNA template is indicated by both crystal complexes (Garcia-Diaz et al., 2005) and simulations (Foley et al., 2006). In Pol*μ* the situation may be similar as in Polλ, as assessed by molecular dynamics simulations (Li and Schlick, 2010).

In addition to large-scale domain motions, several side chains rearrange to assemble the active site. Specifically, the thumb subdomain includes one motif closely implicated in nucleotide selectivity: the side chains of the Tyr and Phe residues forming the YFTGS motif in Polß and Polλ participate in concerted movement upon binding of the incoming nucleotide (Fig. 83B), resulting in minor groove interactions with the base on the primer terminus and correct positioning of the primer terminus for catalysis. This motif is different in Pol*μ* and TdT (GWTGS), where the equivalent residues are a glycine and a tryptophan. The shift in the conformation of this motif observed in Polß and Polλ is unlikely to occur in Pol*μ* or TdT: none of the structures of TdT shows a change in the conformation of these two residues (Delarue et al., 2002). In Pol*μ* and TdT, these Gly and Trp residues are responsible for a decreased discrimination between deoxyribo- versus ribonucleotides (Boule et al., 2001; Nick McElhinny and Ramsden, 2003; Ruiz et al., 2003), as compared to Polß and Polλ, which largely favor the use of deoxyribonucleotides (Beard and Wilson, 1998; Boule et al., 2001; Ruiz et al., 2003; Shevelev et al., 2003). A more general role of this motif during catalysis will be discussed later.

The thumb subdomain is also stablishing interactions with the downstream part of the template strand, through three structural elements: two loops ("nail" and Loop 2) and the already mentioned α-helix N, which is strongly positively charged in Pol*μ*. An interaction common to Pol*μ* and Polλ is the stacking of the templating base with one of the residues included in this positive path, $Arg^{442}$

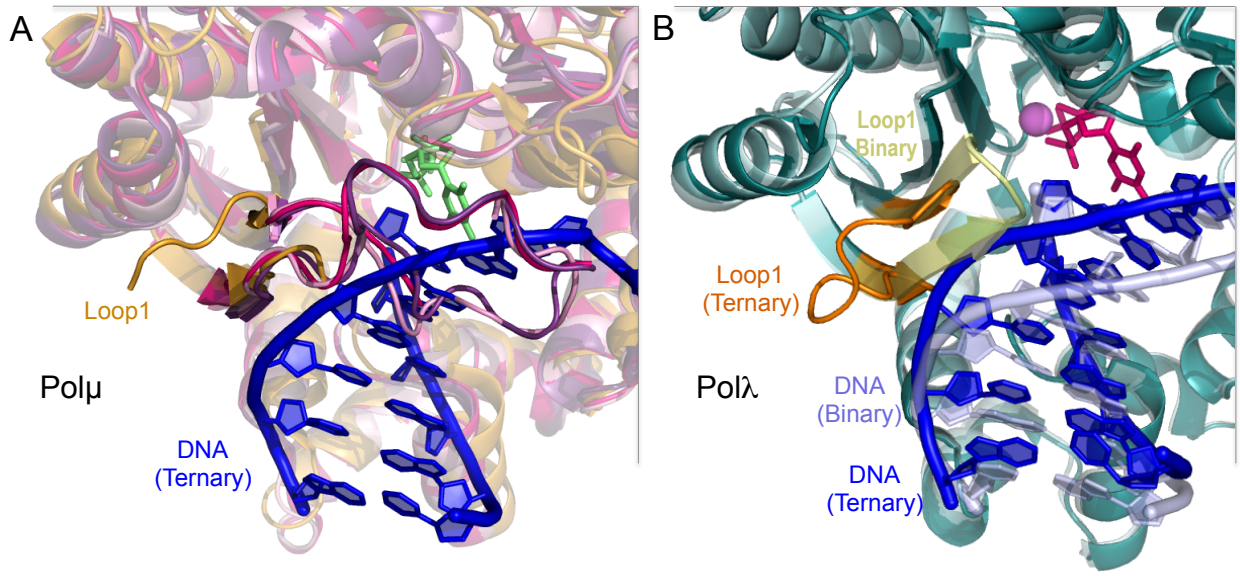



**Figure 84. Loop 1 in Pol*μ*, TdT and Polλ: movements from the binary to the ternary complexes.**
A) Superimposition of the three available crystal structures of TdT (1JMS, 1KDH, 1KEJ, in light pink, dark pink and purple, respectively) and the ternary complex of Pol*μ* (orange). B) Superimposition of the binary (light teal) and ternary (dark teal) complexes of Polλ, shown in cartoon. Loop 1 from both structures are shown in yellow and orange, and the DNA substrates in light and dark blue, respectively.

in Pol*μ* and Arg514 in Polλ (Garcia-Diaz et al., 2005; Moon et al., 2007). The equivalent residue in Polß is a lysine residue (Lys$^{280}$) with a similar function. Pol*μ* has a few minor groove interactions through residue Arg$^{445}$, some of which are mediated by bridging water molecules (Moon et al., 2007). Regarding the loops in the thumb subdomain, and compared to Polλ and Polß, Pol*μ* has a shorter loop in the thumb domain (nail, Asp$^{465}$–Val$^{471}$). This loop in Polλ is suggested to stabilize the extrahelical nucleotide present in frameshift intermediates (Garcia-Diaz and Kunkel, 2006). In Polß, this loop has been found to have an important role in interactions with repair proteins (Gryk et al., 2002). In Pol*μ*, this nail is much shorter than in Polλ or Polß and positioned farther from the binding cleft, a conformation similar to that found in TdT. In the next section of this Discussion, new roles for this motif, and also for the previously uncharacterized Loop 2, will be considered.

## 2. Specific elements conferring the virtuosity.

### 2.1 Why do Pol*μ* and TdT have terminal transferase activity?

Template instruction is a general feature of most members of the X family, with the exception of TdT. TdT is the only known DNA template-independent DNA polymerase, as it is able to add nucleotides to a primer DNA molecule in the absence of a template chain. This feature is crucial for its function in V(D)J recombination, where TdT adds nucleotides to the recombinational junctions of immunoglobulins and TCR receptor genes, generating variability as it creates new information (Bentolila et al., 1995; Bentolila et al., 1999). Interestingly, Pol*μ* shows hybrid biochemical properties: it has an intrinsic terminal transferase activity, but it is strongly activated by a template DNA chain (Dominguez et al., 2000).

Understanding the structural and functional basis of the template-independence of TdT had to await the resolution of the crystal structure of the Polß-like core of TdT (Delarue et al., 2002). A loop region between ß-strands 3 and 4, referred to as Loop 1, has a similar position in all three TdT structures, and is located in a region of the DNA binding cleft that would normally be occupied by the template strand (Fig. 84A). Therefore, this loop could occlude binding of any DNA substrate possessing a template strand, thus explaining its null activity on these substrates. On that basis, and by extrapolation to Pol*μ* of the structural model of TdT, it was predicted that Loop 1, specifically present in these two enzymes, could be directly responsible for their template-independent terminal transferase activity, but in Pol*μ* Loop 1 must be flexible enough to allow template-directed polymerization (Juarez et al.,

2006). In agreement with this prediction, when the crystal structure of Pol*μ* bound to a gapped DNA was solved (Moon et al., 2007), Loop 1 was disordered suggesting conformational flexibility (Fig. 84A). In this structure, the DNA duplex was bound in the usual fashion within the DNA binding cleft. It was then clear that Loop 1 of Pol*μ* cannot occupy the same position as that of TdT when a template strand is present. A comparison of the ends of the ß-strands flanking the loop shows that TdT's Loop 1 extrudes upwards, toward the DNA binding cleft, while that of Pol*μ* appears to turn downwards, away from the cleft (Moon et al., 2007). Although no crystal structure is available of Pol*μ* with a single stranded or 3'-protruding DNA substrate, it is likely that Loop 1 would then be found in the same conformation as in TdT, i.e. interacting with the primer strand, somehow mimicking a template strand. The structural evidence suggested that Loop 1 in Pol*μ* may adopt different conformations depending on the nature of the substrate: the inherent flexibility of this loop in Pol*μ* is distinct from TdT and suggests how Pol*μ* can accommodate different substrates. Studies including the Loop 1 chimeras on Pol*μ* (Juarez et al., 2006) and TdT (Romain et al., 2009) confirmed this hypothesis: replacement of the TdT Loop 1 with that of Pol*μ* is sufficient to allow template-dependent additions, while the reciprocal chimera (Pol*μ* with the TdT Loop 1) is much less inclined to perform template-dependent additions.

The equivalent regions in Polß and Polλ would be less likely to interfere with binding of the template strand because they have a much shorter Loop 1: small enough in Polß to be described as a turn and of intermediate length in Polλ (Fig. 84B). Consistent with this idea, when Loop 1 in Pol*μ* is shortened to a length similar to that of Polλ, the altered polymerase has higher catalytic efficiency on template-containing substrates, but is incapable of template-independent synthesis (Juarez et al., 2006; Nick McElhinny et al., 2005). Consistent with all this, Polλ has strongly reduced ability to catalyze template-independent synthesis, but retains the ability to perform template-instructed additions. Polλ Loop 1 may be involved in a function somehow related to that in Pol*μ*: modulation of fidelity by controlLing dNTP-induced movements of the template strand and 3'-primer terminus in the transition from an inactive to an active conformation of the enzyme (Bebenek et al., 2010). In fact, dNTP binding induces Polλ to transition from an inactive to an active conformation: ß-strands 3 and 4 partially unravel to form Loop 1, a nine-residue loop that repositions as the DNA template strand assumes its active conformation (Fig. 84B). Such a "fidelity checkpoint" would then be related to the energetic penalty of changing the structure of these ß-strands, that would only be overcome in the case of the formation of a correct match.

However, Loop 1 is not the only structural virtue of Pol*μ* and TdT to be able to perform template-independent additions. A second key contribution to template-independent activity comes from a specific histidine residue within the active sites of Pol*μ* and TdT (Moon et al., 2007). This side chain is conserved between Pol*μ* (His$^{329}$) and TdT (His$^{342}$), but is absent in Polß (Gly$^{189}$) or Polλ (Gly$^{426}$) (Fig. 85). In the Pol*μ* ternary structure, the His$^{329}$ side chain adopts a conformation that would allow for hydrogen bonding between the histidine nitrogens and the phosphate of the primer terminal residue

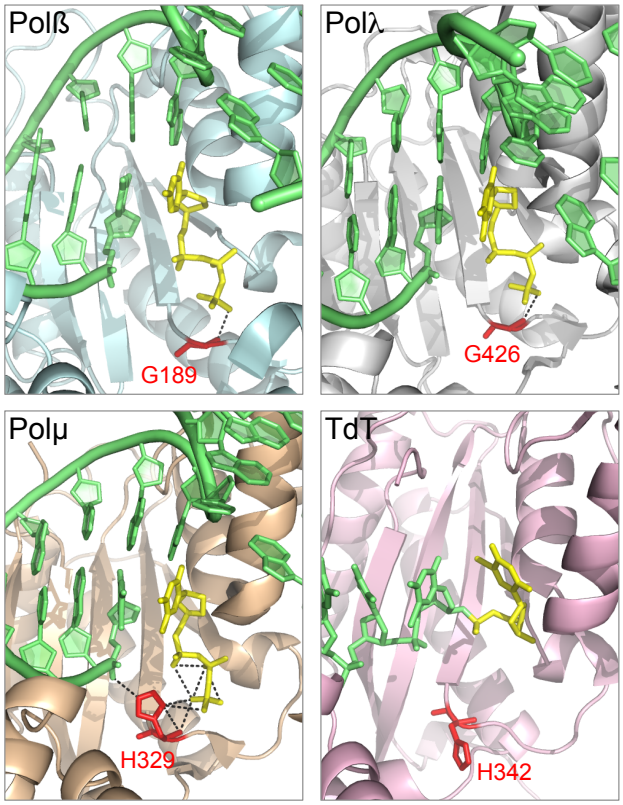



**Figure 85. His$^{329}$ of Pol*μ* and its role as a dual ligand.**
Views of the active site of the X family members: Polß (light blue), TdT (light pink), Pol*μ* (wheat) and Polλ (gray), shown in cartoon. DNA substrate shown in green, incoming dNTP shown in yellow, His$^{329}$ and orthologues shown in red sticks.

and/or the γ-phosphate of the incoming nucleotide. These interactions appear to be crucial for proper positioning of the primer terminus and the incoming nucleotide during template-independent polymerization (Moon et al., 2007). The conformation adopted by Pol*μ* $His^{329}$ is not observed for $His^{342}$ in any of the TdT structures (none representing a ternary complex), but could be easily adopted by a simple rotation of the side chain. This same movement from an inactive to a catalytically competent conformation of the histidine could be also happening during Pol*μ* active site assembly, as indicated by molecular dynamics modelling (Li and Schlick, 2010). Mutating this histidine in Pol*μ* (Andrade et al., 2009; Moon et al., 2007), and TdT (Romain et al., 2009) substantially reduced template-independent activity. In most DNA-dependent DNA polymerases, proper positioning of the 3' terminus is indirectly dictated by the enzyme's avidity by the templating base, thus configuring a binary complex ready to select the incoming nucleotide (ternary complex). Eventually, when no template base is available (blunt or 3'-protruding ends), any further nucleotide addition is unfavored, probably due to a deficient re-positioning of the 3' terminus. The specific role of this histidine during the catalytic cycle of terminal transferase on a single-stranded or 3'-protruding substrate could include relocation of the primer from an unproductive position, as the one adopted by the ssDNA in the TdT cocrystal in which the 3'-OH is occupying the place of the incoming nucleotide (PDB ID: 1KDH), to a productive position, liberating the N-site with a -1 translocation of the primer. This translocation, that has to take place in the first catalytic cycle (due to unproductive binding of the enzyme to the DNA substrate) but also after every round of incorporation, is proposed to be the rate-limiting step in the terminal transferase reaction (Andrade et al., 2009). Thus, is easy to imagine how Loop 1 and the histidine could cooperatively stabilize the primer terminal nucleotide in a catalytically competent conformation, specially when no complementary template strand base is present.

The role of Loop 1 during terminal transferase additions has been now stablished, but a more in depth study of how Pol*μ* fixes and/or orients this mobile part of the protein in accordance with the substrate on which it is polymerizing was performed and accomplished in this Thesis.

Firstly, we made point mutations in the bordering regions of Loop 1, guided by the conservation of those residues at the primary sequence level and by comparison of the available crystal structures of TdT and Pol*μ*: $Phe^{385}$ and $Phe^{389}$. $Phe^{401}$ of TdT (corresponding to $Phe^{385}$ in Pol*μ*), is involved in maintaining the fixed position of Loop 1 *via* a strong stacking interaction between its aromatic ring and $His^{475}$ ($His^{459}$ in Pol*μ*), located in a mini-loop at the thumb subdomain. Mutant F401A in TdT had a striking phenotype, turning a completely template independent enzyme into a DNA-instructed DNA polymerase (Romain et al., 2009). This mutation clearly disrupted the network of interactions needed to maintain a fixed orientation of TdT Loop 1, that is now endowed with a greater degree of flexibility, as in Pol*μ*, thus allowing TdT to accept a template strand. Mutation F385A in Pol*μ*, not affecting templated additions, largely abolished the terminal transferase activity of Pol*μ*, thus confirming that $Phe^{385}$ has a specific role in the catalytic cycle only when a template strand is not available. $Phe^{389}$ is again conserved among Pol*μ*s and TdTs ($Phe^{405}$) of different species, and in both cases it seems to be involved in maintaining the shape and orientation of this motif. Mutation of this residue to alanine in TdT abolishes terminal transferase activity and allows templated insertion of only one nucleotide on a template/primer substrate (Romain et al., 2009). We decided to mutate the equivalent Pol*μ* residue ($Phe^{389}$) to leucine (the aminoacid present in Polλ) and also made the double mutation F389G/R387K. The expected implication of $Phe^{389}$ in the ability of Pol*μ* to catalyze untemplated nucleotide additions was confirmed by testing the terminal transferase activity of these mutants: it was completely abolished, also in the case of the double mutant, something that is specially noticeable since this enzyme also contains the R387K mutation that, alone, boosts this activity by 100-fold.

In a second approach, we mutated the conserved residues (NSH motif) in a mini-loop located in the thumb subdomain. In TdT this loop is interacting with Loop 1 through $His^{475}$, that is conserved in Pol*μ* ($His^{459}$). This mini-loop is also present in the other members of the X family, but its function is different: residues from this loop directly interact with the template strand. In Pol*μ* the role of this loop is an intermediate one: depending on the substrate used and the desired conformation of Loop 1, this mini-

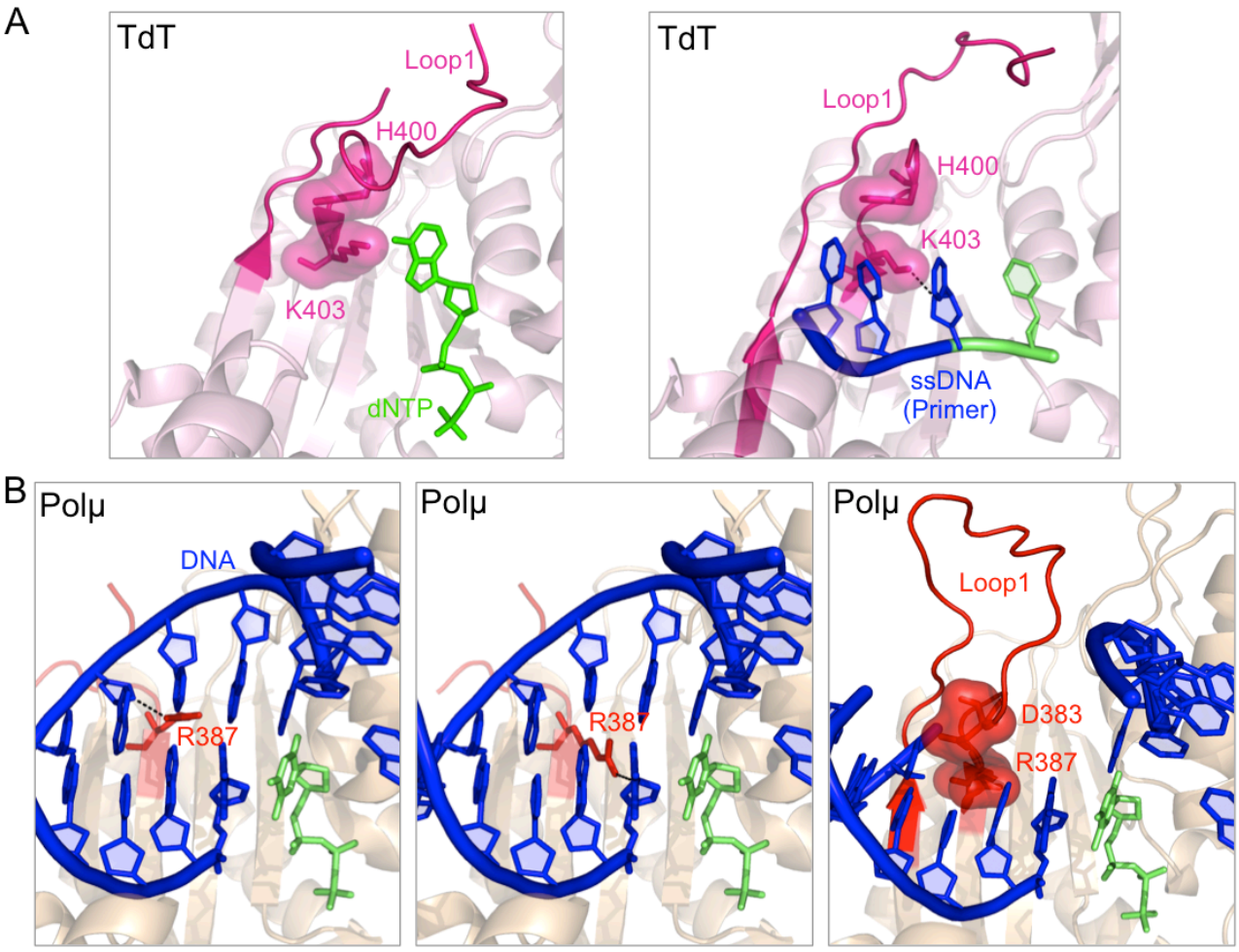



**Figure 86. Arg$^{387}$ triple interactions with the primer and template strands and with Loop 1.**
A) Cartoon representations of the binary complexes of TdT bound to dNTP (1KEJ) or ssDNA (1KDH). Loop 1 is shown in dark pink and Lys$^{403}$ and His$^{400}$ are showin in sticks with semi-transparent surface. B) Cartoon representations of the ternary complex of Pol*μ*: in the left panel, the oriiginal position of Arg$^{387}$ contacting the template strand; in the middle panel, the predicted interaction with the primer; in the right panel, overimposition of Loop 1 from the TdT structure (1JMS) and proposed interaction of Arg$^{387}$ and Asp$^{383}$.

loop may interact with the template strand (through Asn$^{457}$) or with Loop 1 (through His$^{459}$). Accordingly, the asparagine is only needed during templated additions, and dispensable for terminal transferase λactivity of Pol*μ*, while the histidine had the opposite effect. We propose a regulatory function for the NSH motif in the thumb mini-loop, helping to accommodate either the template strand (as in Polß of Polλ) or Loop 1 (as in TdT) as suits best for each individual situation.

Having now a general idea of how this two polymerases, Pol*μ* and TdT, are specially designed to perform this untemplated additions of nucleotide units, another question still remains: why the terminal transferase activity of TdT is much higher than that of Pol*μ*? This topic has been the subject of examination and has been answered in this Thesis. Combined structural and functional evidences for both Pol*μ* and TdT, together with the functional data presented in this Thesis, indicate that there is another residue directly implicated in the terminal transferase activity of both enzymes. That residue (Arg$^{387}$ in Pol*μ* and Lys$^{403}$ in TdT) modulates the catalytic efficiency of the terminal transferase reaction, by regulating the rate-limiting step. Judging by the structural data available, this residue could be stablishing dual and alternative interactions during the catalytic cycle of both Pol*μ* and TdT: when the primer is bound at the unproductive position (TdT crystal 1KDH), the residue is interacting with the primer strand, while in the Pol*μ* crystal in which the primer strand is correctly positioned in a productive complex (2IHM), the arginine is interacting with the -3 position of the template strand. In the case of Pol*μ*, and assuming an alternative interaction as that seen in TdT, Arg$^{387}$ acts as a brake for the necessary

movement of the primer, to limit nucleotide additions before end bridging. In fact, the single change of this residue for the TdT counterpart (Pol*μ* mutant R387K) showed an increase in untemplated additions that ranged from 10- to 100-fold, reaching levels comparable to those of TdT itself (this PhD Thesis, {Andrade et al., 2009). Interestingly, mutant R387K produced a very specific blockage at position +4 when continuous terminal transferase extension of a blunt end was tested (this PhD work, {Andrade et al., 2009). This situation is such that in a 3-protrusion of 4 nt the second proposed protein-DNA interaction for this residue cannot occur, since the -3 position of the template strand is not available. In these substrates (ssDNA, 3' protrusions longer than 3 nts), this residue must be adopting a new partner for this second interaction, most surely a portion of the protein that is now located in place of the template strand: Loop 1. TdT Loop 1 contains a histidine ($His^{400}$) that completely superimposes with the -3 position of the template strand, and this histidine is surely acting as a partner for $Lys^{403}$ when it is not interacting with the primer (catalytically active configuration; Fig. 86A. left panel). In agreement with this, our results measuring TdT activity on substrates ranging from blunt to 11 nt 3'-protruding indicate that polymerization was inhibited when the protrusion was shorter than 3 nt (these substrates would not allow correct positioning of Loop 1 and $His^{400}$). A similar protein-protein interaction between $Arg^{387}$ and Loop 1 is surely occurring in Pol*μ* when the -3 position of the template is not available (Fig. 86B, right panel), and it is distorted when the arginine is mutated to alanine, as indicated by the completely defective terminal transferase activity of mutant R387A.

Interestingly, the equivalent residue in human Polλ ($Lys^{472}$) is also involved in regulating the catalytic cycle by means of inhibitory interactions with the primer strand (Bebenek et al., 2010). Recent results suggest that $Lys^{472}$ may help to modulate template-dependent synthesis. In the wild type Polλ binary complex (1XSL), $Lys^{472}$ is within H-bonding distance of the 3'-O of the primer terminal nucleotide. Such hydrogen bond between $Lys^{472}$ and the primer terminus that could stabilize the inactive conformation must be disrupted in order for the 3'-O to assume its catalytically competent position. A weakened interaction between $Lys^{472}$ and the primer terminus would allow the 3'-O to more easily adopt a conformation that would support catalysis with an incorrect nucleotide bound, reducing the discrimination between correct and incorrect incorporation (Bebenek et al., 2003).

Thus, $Arg^{387}$ plays a key role in modulating template-independent synthesis by Pol*μ*, having a dual role: it allows terminal transferase additions to occur, but also acts as a brake that limits these additions. Substituting the homologous lysine in TdT with arginine or alanine (Romain et al., 2009) also results in loss of template-independent activity, although the properties of the two TdT mutants are not identical. In the case of TdT, residue $Lys^{403}$ likely establishes a weaker interaction with the primer compared to its orthologue $Arg^{387}$ in Pol*μ*. Thus, TdT has been optimized to efficiently overcome the rate-limiting step of the terminal transferase, to exclusively perform creative synthesis.

What is the reason for this limited terminal transferase activity in Pol*μ*? Our results indicate that when a templating base is provided *in trans* during NHEJ, the rate-limiting step is relieved. A templating base provided *in trans* by the approaching end that could be located in a proper register will stabilize the incoming (and complementary) nucleotide, thus facilitating primer translocation. As a result of this, NHEJ of many incompatible ends can be efficient and accurate. During NHEJ of this fraction of incompatible ends, an excessive terminal transferase as that displayed by mutant R387K would be disadvantageous in terms of genome stability. On the other hand, our findings also explain the need for a mild terminal transferase activity in Pol*μ*, not only to create connectivity in those other DNA ends that cannot be efficiently joined on a templating basis, but perhaps contributing to gain a certain degree of genome variability. Additionally, it can be inferred that TdT evolved to maximize the efficiency of the translocation mechanism in the absence of template, at the cost/benefit of introducing untemplated nucleotides, thus being devoted to generate variability at V(D)J recombination intermediates.

Is this the physiological role of the terminal transferase activity of Pol*μ*? As shown here, and in other reported studies, NHEJ of short incompatible ends can be accurate in many cases, but imprecise in others depending on both the length and sequence of each protrusion. For the latter cases, when a templating base is not in a proper register,

untemplated terminal transferase addition in a NHEJ context provides a valid, although mutagenic, solution that would be conceptually similar to translesion DNA synthesis. Besides, it cannot be ruled out that Pol*μ*'s terminal transferase can extend a single short 3'-protrusion to facilitate end-joining of this fraction of non-complementary ends.

On the other hand, there is also *in vivo* evidence of untemplated insertions made by Pol*μ*. It has been shown that mice that are TdT-/- still contain a 5% of their V(D)J junctions with template-independent additions, which suggested a possible role of Pol*μ* in these reactions (Bertocci et al., 2006). In agreement with that, the terminal transferase activity of Pol*μ* has been directly implicated in variability/repair processes occurring at embryo developmental stages in which TdT is still not expressed (Gozalbo-Lopez et al., 2009).

**2.2 Pol*μ* in NHEJ: The dexterity of clutching two chains with one hand.**

The NHEJ pathway takes advantage of highly specialized polymerases, capable of dealing with 2 DNA ends at once: one providing a broken primer strand with a protruding 3'-OH, and a second end that provides a recessive 5'-P, and a broken template strand, also 3'-protruding. Thus, for this task, two different pieces of DNA need to be hold by the same enzyme molecule. A great amount of evidence indicates that the NHEJ pathway minimizes loss of genetic material by using any template available: the overhangs formed during alignment of DSBs are usually filled in, allowing retention of the maximum sequence possible (Della et al., 2004; Kramer et al., 1994; Roth and Wilson, 1986; Thode et al., 1990). In this context, a NHEJ polymerase must use a primer provided by one end and a template provided *in trans* by a second end: "alignment based gap fill-in". In the case of family X polymerases, the opportunity to act during NHEJ is dictated by their template preference, a property that follows a gradient ranging from Polß, that only polymerizes on substrates with a continuous template strand; to Polλ, that is active in NHEJ only when the template strand is stabilized by complementarity with the primer strand; to Pol*μ*, who can direct template-dependent synthesis even when there is no base-pairing between the two ends; to TdT, that also acts on unpaired primer termini but does not allow the use of a template strand (reviewed in Ramsden, 2011). It has been suggested that this variable degree of template dependence relies on structural differences among the four polymerases. In this section we will decipher these structural determinants, in particular those conferring Pol*μ* the unique handiness of trans-polymerizing without the help of a single base pair.

The first thing that comes to mind when studying NHEJ-specific polymerases is the common presence of the BRCT domain. As already mentioned, this domain is involved primarily in mediating protein-protein interactions that recruit the DNA repair enzymes to the DNA break. We have shown that Pol*μ* is able to join two DNA ends in the absence of any NHEJ accessory factors such as Ku or XRCC4/LigIV. On the other hand, the BRCT domain was crucial when testing the end-joining capacity of Pol*μ* and Polλ in the presence of NHEJ factors. Taken together, this evidence suggests that even though it is likely that when classical NHEJ takes place the polymerases need to be recruited through their BRCT domains, Pol*μ* could participate in the alternative end-joining pathway that takes place without the need for the classic NHEJ factors due to its intrinsic avidity for DNA ends. In fact, our results indicate that the BRCT domain of Pol*μ* does not play a critical role in the catalytic step of the NHEJ reaction, but could benefit NHEJ by promoting a tighter binding of Pol*μ* to the template-providing DNA end.

Looking more closely at the polymerase domain, a new motif at the very N terminal portion of Pol*μ*, TdT and Polλ was found, that acts as a brooch that holds together the 8 kDa domain and the thumb subdomain. This motif is highly conserved among Pol*μ*s and TdTs from very different species, being present in *S. pombe Sp*Pol4, and is formed by the five amino-acids YACQR. Similar residues are conserved in the sequence of PolXs from many *Ascomycota*. In the case of Polλ the motif is different (WxCxQ) but it is equally conserved. *Sc*Pol4 from *S. cerevisiae*, strikingly, lacks this motif, the same as Polß, where the polymerase core starts just in the middle of this motif, without any sequence conservation in this zone. Taking into account the structural data available we could initially postulate the implication of this YACQR motif in maintaining the permanent "closed" conformation of Pol*μ*, TdT and Polλ through the catalytic cycle, while Polß, that

is not constrained by this “brooch”, allows free motion of the 8 kDa domain with respect to the thumb, and adopts an “open” conformation when no substrate is available. This level of flexibility inversely correlate with the level of implication of each polymerase in "difficult" reactions such as those involving a compromised continuity of the template strand, that take place during NHEJ. In these reactions, a tightly controlled conformation of the body of the polymerase would help to correctly position all the components needed for catalysis: the primer terminus, the templating base, the incoming nucleotide and the protein side chains building the active site. In agreement with this, our mutagenesis results indicate that the most affected reaction when this motif is disturbed is bridging of non-compatible ends by Pol*μ*, as these substrates do not offer any additional stabilizing potential strength to that provided by the enzyme itself for end alignment. On the other hand, the support provided by the “brooch” was trivial during gap-filling reactions. The same conclusion applies to Polß, a polymerase specialized in gap-filling that has probably lost the “closing” motif to gain other beneficial abilities that will be described more in depth in the last section of this Discussion. In the case of Polλ, it remains an open question if the counterpart of the “brooch” is still functional or if this enzyme is more ß-like in this regard, since our results with Pol*μ* indicate that the importance of the maintenance of this 8 kDa-thumb interface is less critical when dealing with compatible substrates as those required by Polλ during NHEJ.

Regarding binding to the NHEJ substrates, still more Pol*μ*-specific interactions were found and studied during the course of this PhD work, and those can be clearly divided into two surfaces dedicated to each of the substrates that need to be joined: 1) one platform is devoted to binding of the downstream part of the template strand, including the templating base, through an arginine-rich helix that stands just beside the phosphate backbone of this part of the substrate; 2) a second interaction surface whose duty is to bind and aid the approach of the primer strand. These two regions have different levels of conservation in the four members of the X family, but the whole cohort of positive charges in the primer-binding zone are only present in Pol*μ* and TdT, while the four arginines that shape the template-binding zone are only conserved in Pol*μ*. More specifically, the primer binding-zone has been studied here by mutations in $Lys^{249}$, $Arg^{253}$, $Arg^{416}$ and $His^{329}$. All of these mutations affected specifically the reactions in which a template strand is discontinuous: terminal transferase and end-joining, while not affecting 1 nt gap-filling at all (except for the case of $Arg^{416}$, which is conserved in all X family members and directly contacts the 3’-OH of the primer). These residues proved necessary only in the case of polymerization on longer gaps, where binding to the 5’-P implicates that the primer terminus will be located further away from the active site (as shown in the cocrystal of the binary complex of Polλ with a 2nt gap). These residues will be needed in this situation to “pull” the primer strand and bring it closer to the 5’-P, while at the same time producing a “bubble” in the template strand that will be stabilized in a different manner by each polymerase (this topic will be extensively discussed later). The level of conservation of these residues, being all of them present only in Pol*μ* while Polλ and Polß have lost some of these contacts, alludes to their importance for Pol*μ* specific activities that are not present in the other enzymes, namely the ability to carry out untemplated additions and also the exceptional talent of bridging and execute template-directed insertions on two non-complementary DNA ends.

Regarding the downstream template binding, Pol*μ* has an extraordinarily positively charged helix that holds the phosphate backbone due to the presence of four arginine residues. Of this four positive charges, only two are conserved in Polλ or Polß, and their role is related with fidelity through their interaction with the template of the nascent base pair. In Polλ, these residues also are involved in controlling the motion of the thumb occurring during the transition from the binary (E:DNA) to the ternary (E:DNA:dNTP) complex, that has the pursued effect of bringing the template strand closer to the thumb into its final catalytic position. A proposal that these residues play a similar role in Pol*μ* was initially supported by molecular dynamics simulations (Li and Schlick, 2010). Our results further confirm this hypothesis, both using substrates with continuous template such as DNA gaps, that are correctly configured on their own, but specifically and more drastically during stabilization of two DNA ends during NHEJ. Our results indicate that these arginines do not contribute too much to general binding to the downstream end, given that the presence of the 5’-P group provides most of the

binding strength, but most likely these arginines are implicated in the bridging and positioning of the DNA substrates in a proper register. Based on our results with single stranded substrates we are able to propose a second function for this arginine-containing helix: through indirect interactions, mediated by the thumb mini-loop, that could be improving the stabilization of Loop 1 in the case of maximal closure, i.e. during terminal transferase addition of nucleotides, as well as during NHEJ reactions in which the involvement of this motif could be dual, mediating protein-protein as well as protein-DNA interactions. This idea is further supported by new rounds of molecular dynamics in which these residues were individually mutated to alanine (Li and Schlick, 2010): this simulations, in the first place, displayed a conformation of the template strand similar to that observed for the binary complex in Polλ, demonstrating *in silico* the effect of these positive charges in pulling the downstream part of the substrate to its correct position; secondly, the final distance between Loop 1 (modeled for the simulations) and the arginine-helix is unusually large, and the former is in a conformation that may hamper its interactions with the DNA substrate. Thus, having already discussed the roles of Loop 1, the thumb mini-loop and the arginine-helix, during untemplated additions and also during catalysis on discontinuous substrates, it seems now clear that these three portions of Pol*μ* work coordinately. In Pol*μ*, Loop 1 is considered to be a highly flexible piece that could eventually adopt an indefinite number of conformations, but we argue that this is not so, but that Loop 1 is configured in the apoenzyme as found in TdT, kept in place by interactions with the thumb mini-loop. We propose that Loop 1 would maintain this position in a binary complex in which the DNA substrate is single stranded or contains long (+3 nt) 3'-protrusions and no 5'-phosphate group. In the case of binding a substrate containing a continuous template strand (i.e. a gap), the Loop 1 would become disordered due to its flexibility, provided by its aminoacid sequence and probably also by the labile nature of the interactions keeping it in place. A similar transition occurs even in the case of the mini Loop 1 present in Polλ: as dNTP binding induces the transition from binary to a ternary complex, ß-strands ß3 and ß4 partially unravel to form Loop 1, a nine-residue loop that repositions as the DNA template strand assumes its active conformation. In fact, one of the initial ß-strands is occupying the path that is filled by the template strand in the ternary complex. Finally, when Pol*μ* binds two 3'-protruding NHEJ substrates, Loop 1 needs to adopt a third conformation of inherent flexibility, to accommodate and stabilize the several possible locations and lengths of the gap formed after synapsis in the template strand. In this case, direct protein-protein interactions with the thumb mini-loop and indirect ones with the arginine-helix would assist Loop 1 to be positioned in the correct orientation needed in each case.

The degree of variability at the DNA substrate level during NHEJ compels Pol*μ* to have a spacious and versatile active site. In fact, our results indicate that this versatility is such that Pol*μ* can take advantage of the widest range of solutions when polymerizing: not only it can use deoxy- and dideoxy-nucleotides, but also ribonucleotides (Nick McElhinny and Ramsden, 2003; Ruiz et al., 2003). This lack of sugar discrimination is due to the change of the YF motif, the "steric gate" conserved in Polß and Polλ active sites, for a GW motif. Specifically, the change of the tyrosine for a glycine enables Pol*μ* to use NTPs during catalysis. Recent results of directed mutagenesis in Polλ (Brown et al., 2010) and Polß (Cavanaugh et al., 2011) confirmed the initial results in Pol*μ*, consolidating the idea of a common mechanism of sugar selection in this family of polymerases. In the case of Pol*μ*, this selectivity is lost in the way to find a more open architecture for the active site, that not only allows the insertion of ribonucleotides, but surely contributes also to the low fidelity levels of the enzyme. Thus, Pol*μ* has to cope with both "collateral damages", in order to obtain the profit of being able to deal with non-complementary DNA end substrates. But the ability to insert NTPs brought its own benefits: NTPs are a much more abundant resource than dNTPs during almost all the phases of the cell cycle, something that qualify NTPs as favorable substrates in the equation of DNA repair. Thus, X family DNA polymerases that exhibit low ribonucleotide discrimination would be predicted to insert more NTPs than dNTPs when the *in vivo* dNTP/NTP imbalance is taken into consideration. Also, the use of NTPs during NHEJ *in vivo* is supported by the finding that a terminal ribonucleotide stimulates the activity of eukaryotic ligase IV (Nick McElhinny and

Ramsden, 2003). Moreover, the ability of Pol*μ* to extend with NTPs is dramatically reduced with each successive ribonucleotide added, effectively limiting synthesis to fewer than five nucleotides (Nick McElhinny and Ramsden, 2003; Ruiz et al., 2003). This may be helpful in curtailing template-independent activity of these enzymes, because long non-complementary tails would be difficult to align and ligate. Strikingly, our results show that during NHEJ the error-proneness of Pol*μ* was lowered when using NTPs. This was valid even in the presence of manganese, an optional activating metal ion that appears to be physiological during NHEJ. Mn2+ was for a long time considered mutagenic, and consequently not physiological, but renewed efforts in the study of polymerase function *in vitro* has led to demonstrate that some specialized polymerases, as Polλ clearly prefers to utilize $Mn^{2+}$ even when $Mg^{2+}$ is in large molar excess (Frank and Woodgate, 2007). Our results with Pol*μ* point to a similar conclusion: although able to use other cofactors such as zinc or cobalt, the peak of polymerase activity with NHEJ substrates was reached at the physiological concentration of manganese ions, while this same situation for magnesium was inhibitory. When NHEJ reactions were performed at this low concentration of manganese, the well known burst in efficiency was achieved, but without the negative effect on fidelity that can be observed at the exceedingly high concentrations of manganese used before in the literature. As already mentioned, these positive effects on both efficiency and fidelity were observed not only with dNTPs, but more interestingly, even better results were obtained with NTPs. Whether a similar situation occurs *in vivo* is an intriguing possibility certainly worth considering. By using wild-type HeLa nuclear extracts, and yeast extracts over-expressing the NHEJ related polymerases *Sp*Pol4, Pol*μ* and Polλ, we tested ribonucleotide usage by these enzymes during the first step of a NHEJ reaction and also during the second step (gap-filling), and evidence shows that under these conditions, *Sp*Pol4 and Pol*μ* were able to use NTPs as competent substrates.

Taken together, all these different strategies and inventiveness found in a small enzyme, engineering several protein motifs that held the DNA or other protein pieces of the polymerase in place for catalysis, and also the ambidexterity of the active site, prepared for almost any kind of DNA and nucleotide substrates as well as cofactors that need to be used, speak of Pol*μ* as a highly specialized polymerase, singularly suited for its function: engagement of two different DNA chains and trans-directed polymerization, even in the case of inexistent complementarity, with the only objective of fulfilling the most efficient reaction possible, with the minimal loss of sequence, during NHEJ.

### 2.3 Template selectivity: the modular architecture of X family polymerases.

Structures of X family polymerases resolved over the past 10 years have provided a vast amount of information about how the members of this family deal with the fidelity issue of correctly selecting the templating base in different situations. Mutations can arise as a result of single-base substitutions or as insertions/deletions (indels) of nucleotides during polymerization. The four human enzymes display very different phenotypes in terms of fidelity: Polß has the highest fidelity; Polλ is less accurate for base substitutions and much worse at single-base deletions (Bebenek et al., 2003; Picher et al., 2006); Pol*μ* is highly error prone for frameshifts (Zhang et al., 2001) and substitutions via transient misalignment (Ruiz et al., 2004; Tippin et al., 2004); TdT, being template-independent, cannot really account for any fidelity measurement. X family enzymes do not possess a 3'=>5' exonuclease activity to proofread errors, and have much less contacts with the DNA minor groove upstream the active site than replicative polymerases, a characteristic that probably makes them more prone to accept DNA distortions upstream the active site. These distortions are common during NHEJ, and Pol*μ* (Covo et al., 2004; Duvauchelle et al., 2002; Havener et al., 2003; Zhang et al., 2002) and Polλ (Belousova et al., 2010; Blanca et al., 2004; Crespan et al., 2007; Crespan et al., 2007; Maga et al., 2007; Picher and Blanco, 2007; Zhou et al., 2008) sustain significant activity either when the templating nucleotide is damaged or when base pairs near the primer terminus include damaged nucleotides.

In terms of single-base substitution errors, X family polymerases frequently confront a situation with a highly mutagenic outcome: one of the most common oxidative DNA lesion, 8oxoG, can be present during

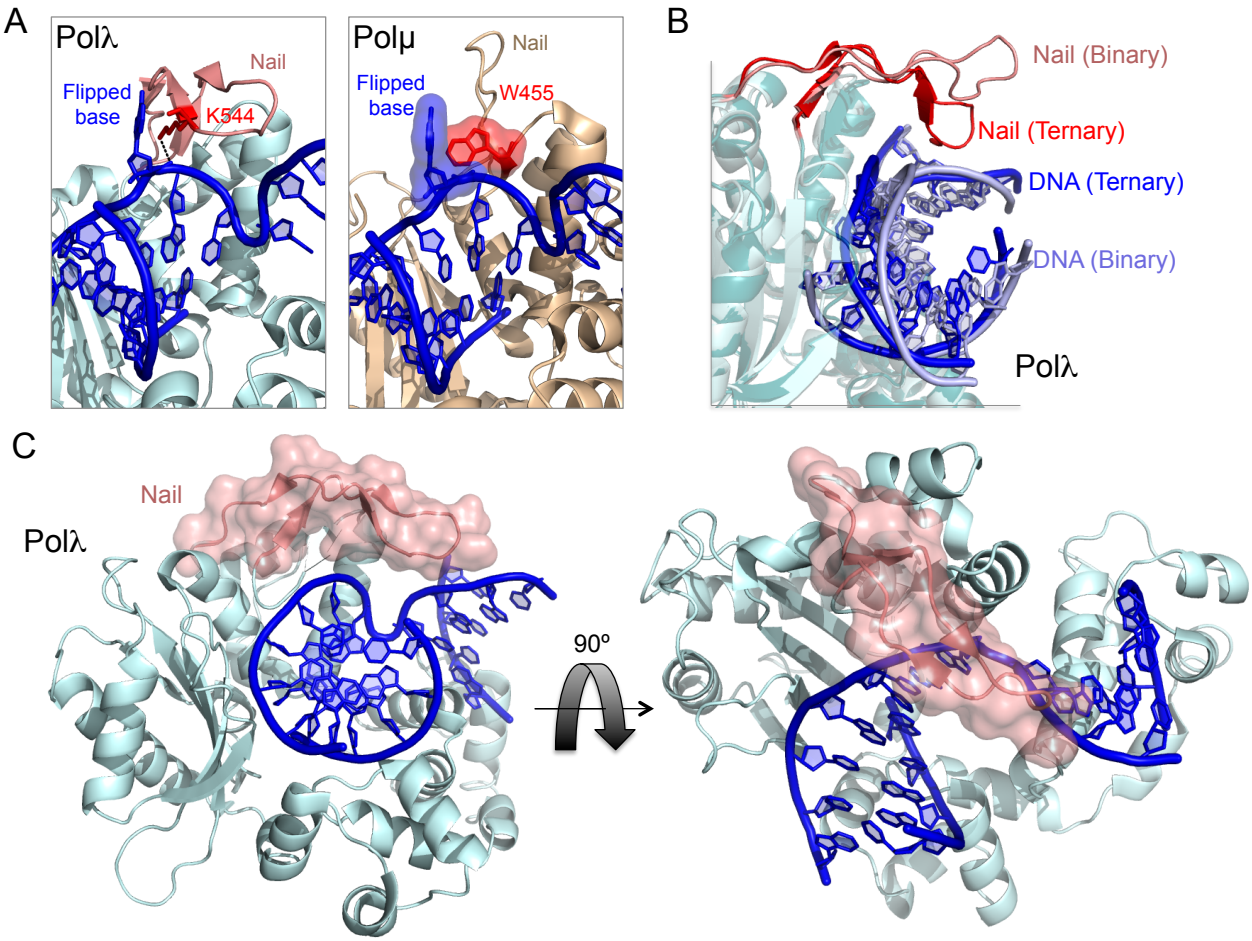
BER-associated reactions and also during NHEJ (Ward, 1981), and the polymerases implicated in both repair pathways must be able to deal with this modified base. Unmodified deoxyguanine prefers an *anti* glycosidic conformation, whereas 8oxoG favors a *syn* conformation that can form a Hoogsteen base pair with adenine. Due to ambivalent base pairing properties, DNA polymerases exhibit low fidelity when encountering 8oxoG as template, inserting dCTP or dATP. Polß has low accuracy and also a low relative efficiency when using damaged versus undamaged bases (Miller et al., 2000). Although human Polλ possesses a high relative efficiency for inserting nucleotides opposite 8oxoG, the accuracy is again low, giving rise to the insertion of dA with a high probability (Picher and Blanco, 2007). In the case of Pol*μ*, fidelity is even lower than that of Polß or Polλ: it only inserts dA in front of an 8oxoG either in a template/primer, gap or, most importantly, during NHEJ. Also, it preferentially extends the 8oxoG:dA mismatch *versus* the correct 8oxoG:dC match, when present in the connection of two DNA ends. The efficiency of Pol*μ* in these reactions involving insertion in front of 8oxoG or extension of the damaged or undamaged pairs, was comparable to that of normal polymerization. The structure of Polß with dCTP and a DNA gap where the templating base is 8oxoG (Krahn et al., 2003) illustrates that phosphate-backbone flexibility in the template strand is a critical parameter that can modulate the *anti*/*syn* equilibrium of the templating nucleotide. A close examination of the structure of the ternary complex of Pol*μ* leads us to propose that the large number of positive charges surrounding the templating base (arginine-helix) restrain the number of conformations that can be adopted by the DNA phosphate backbone in this region. As already discussed, this is of enormous interest in order to reinforce the position of the template in the case of joining two non-complementary ends (this PhD thesis), but can prove unbeneficial for the bypass of this specific damage, since all these interactions are most probably keeping 8oxoG in the mutagenic *syn* conformation when acting as template. In any case, the ability of Pol*μ* to make use of 8oxoG with such efficiency would be beneficial if this form of oxidative damage is present in the NHEJ synapsis, perhaps facilitating connectivity of non-complementary ends.

Indels are common errors produced during DNA replication and repair, associated with different human pathologies including cancer and diseases associated with expansion of repeats. A one- or two-nucleotide indel in a coding sequence results in a shift in the reading frame (frameshift mutation) that can abolish the activity of the protein. The first hypothesis explanatory of the production of indels was introduced by Streisinger et al., 1966: this fameshift mutations were described as products of strand slippage in repetitive DNA sequences, and after the formation of misaligned intermediates, the deletion or insertion of unpaired bases would occur. Other two models have been proposed since then, namely direct misincorporation misalignment, when a polymerase introduces an initial mismatch that causes the primer terminus to be subsequently realigned (Bebenek and Kunkel, 1990; Kunkel, 1988) and dNTP-stabilized misalignment, when incorporation of the correct dNTP occurs in front of a complementary downstream template base (Bloom et al., 1997; Efrati et al., 1997). All polymerases studied to date generate indels during DNA synthesis *in vitro* (Kunkel, 2004), but with very different frequency. X family members Polß, Polλ, and Pol*μ* all generate single-base deletions during synthesis (Bebenek et al., 2003; Kunkel, 1985; Zhang et al., 2001). Polλ generates single-base deletions at a much higher rate than Polß (Bebenek et al., 2003), a feature that may correlate with the fact that Polλ has much less contacts with the template strand upstream the polymerization site. On top of that, crystal structures of Polλ with misaligned frameshift intermediates (Garcia-Diaz and Kunkel, 2006) showed that a loop located close to ß-strand 8 in the thumb, that we refer to as "nail" (as a natural extension of the thumb), is interacting with the "looped-out" nucleotide through a positively charged residue, $Lys^{544}$ (Fig. 87A, left panel). This contact with the extrahelical nucleotide seems to direct the position of the misalignment one base pair upstream of the polymerase active site, even if the substrates were designed to favor positioning of this distortion in further upstream locations. $Lys^{544}$, together with $Arg^{538}$ located also in the nail, are also stablishing direct interactions with the -1 and 0 positions of the template strand, while the rest of the residues in this long loop are "ironing" the downstream part of the template strand (Fig. 87C). These features are ideal for a polymerase whose relevant substrates may have imperfections in the duplex upstream of the polymerase active site,

and thus may be critical for the efficiency of the end-joining reaction. Pol*μ* is also capable of deletion synthesis (Tippin et al., 2004; Zhang et al., 2001), and the resulting intermediates may be stabilized in a similar fashion. The stabilizing lysine from Polλ, $Lys^{544}$, is not conserved in Pol*μ*, but an alternate residue, $Trp^{455}$, could perform a similar role (Moon et al., 2007; Fig. 87A, right panel). When overimposing the Polλ misaligned DNA intermediate on the Pol*μ* ternary structure, this residue is located very close to the flipped-out base, and could stabilize it through stacking interactions.

The interactions of the nail with the phosphate backbone of the DNA substrate are conserved in the ternary complex of Polλ with a 1 nt gap in which the substrate is correctly aligned, and this leads us to propose that these interactions may serve a more primary function. The nail is not a rigid piece during the catalytic cycle, but it changes its position from the binary to the ternary complex at the same time as the DNA substrate moves closer into the active site. These two movements seem to be correlative, and the shift in the position of the nail seems to "push" the phosphate backbone of the template strand into its final catalytic position (Fig. 87B). Accordingly, a deletion mutant of Polλ lacking the nail displayed much lower catalytic efficiency than the wild-type enzyme (Garcia-Diaz and Kunkel, 2006). Also, the nail might serve to help producing a very different type of distortions in the DNA. As already explained, there is a structurally imposed difficulty for these polymerases when filling a gap longer than 1 nt: binding to the 5' end of the gap through the phosphate, and the obligatory positioning of the 3'-primer terminus close to the nucleotide site, imposes that the template strand

**Figure 87. Role of the nail in the generation of frameshifts.**
A) Cartoon representation of Polλ (light teal) and Pol*μ* (wheat) showing the DNA substrate in dark blue sticks, the nail in pink (in Polλ) and selected residues in red sticks, B) Superimposition of the Polλ binary and ternary complexes, with the nail colored pink and red and the DNA substrate colored light and dark blue, respectively. C) Crystal structure of Polλ showing the semi-transparent surface of the nail and its position with respect to the template strand (dark blue sticks).



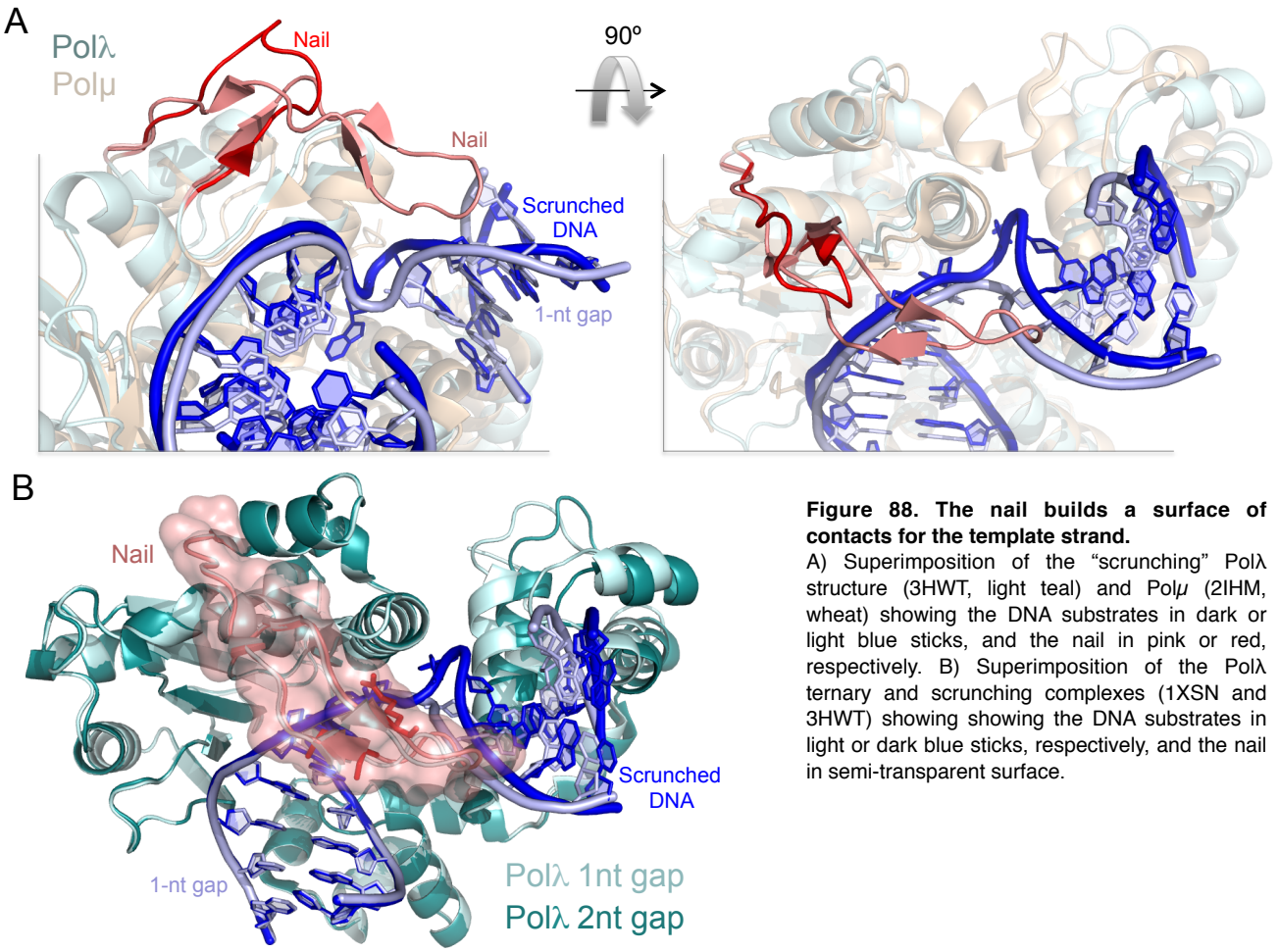



**Figure 88. The nail builds a surface of contacts for the template strand.**
A) Superimposition of the "scrunching" Polλ structure (3HWT, light teal) and Pol*μ* (2IHM, wheat) showing the DNA substrates in dark or light blue sticks, and the nail in pink or red, respectively. B) Superimposition of the Polλ ternary and scrunching complexes (1XSN and 3HWT) showing showing the DNA substrates in light or dark blue sticks, respectively, and the nail in semi-transparent surface.

must form a bubble when it's longer than 1 nt. A recently solved structure of Polλ showing a ternary complex with a 2 nt gap and the nucleotide complementary to the first template in the gap (Garcia-Diaz et al., 2009), proved that necessity, showing that the second template nucleotide in the gap is "scrunched" in a pocket located between the thumb and 8 kDa domains, and stabilized through contacts with 3 residues in that pocket (Fig. 88B). An "opening" movement of the 8 kDa domain is also observable and might help to accommodate the yet-to-be-copied nucleotide (Fig. 88B). When this "scrunched" substrate was overimposed on the available Pol*μ* structure, a similar pocket was apparent, in which some of the residues are conserved and within polar-contact distance to the second nucleotide in the gap (Fig. 88A). Although all the structure determinants seem to be at least partially conserved, Pol*μ* does not behave like Polλ during polymerization in a long gap: while Polλ (and Polß) fills the gap step by step, correctly "counting" and copying the total number of available templating bases in the gap (Garcia-Diaz et al., 2002), Pol*μ* tends to produce deletions by inserting the nucleotide complementary to the last templating base, i.e. the one closest to the downstream 5'-P (this PhD Thesis; Juarez, 2006). This very different behaviour indicates that there must be something else helping Polλ and Polß to "count", i.e. to correctly "scrunch" the remaining bases while copying the one closest to the primer-terminus. We propose that this assistance is provided by the long nail, conserved in Polλ and Polß, but not in Pol*μ* or TdT. The nail would help "pushing" the +2, +3, etc. templating nucleotides into the scrunching pocket by "ironing" the template strand. In the case of Pol*μ*, due to the lack of a long nail, the template strand could more easily be scrunched in the opposite direction (template misalignment), allowing the last template positions in the gap to be efficiently copied, thus favoring the characteristic Pol*μ*-generated frameshifts (Zhang et al., 2001). In fact, Pol*μ* is different from Polß in the frameshift efficiency (Tippin et al., 2004): on reiterative sequences of only 2 nt

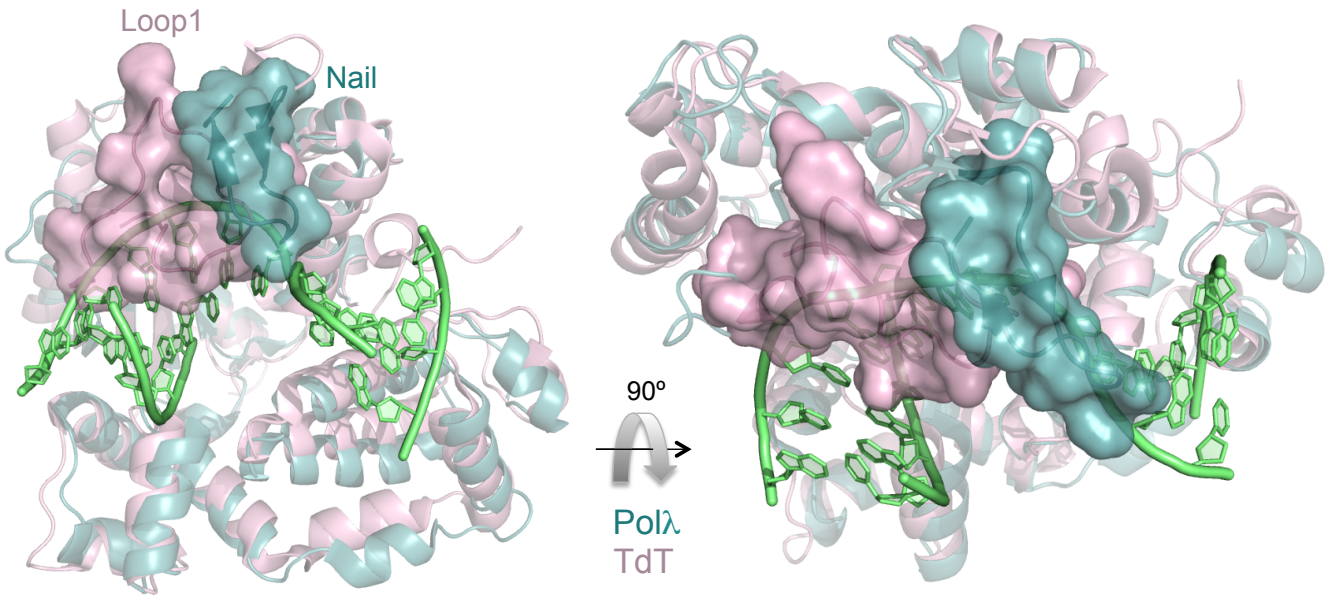
Pol*μ* already produces deletions, while Polß needs longer repetitions *in* order to show slippage propensity. The authors also show that Pol*μ* preferentially produces frameshifts through the Streisinger's "slippage" mechanism, but the "dNTP-selection" mechanism is also utilized in order to "skip" the first templating base (see also Ruiz et al., 2004).

Studying the primary sequence of family X polymerases, including the ones present in vertebrates, yeast, and the X family polymerases predicted in other fungi, we found out that the two main loops in these polymerases involved in interactions with the template (Loop 1 and nail) seem to be incompatible with each other: enzymes that contain a long nail always bear a short Loop 1, and vice versa. Moreover, when overimposing the atomic structures of nail-containing polymerases and Loop 1-containing polymerases, it is striking to notice that these two loops occupy similar positions and clash with each other (Fig. 89). Therefore, our prediction is that the functions provided by these two modules are incompatible and during evolution, each new polymerase had to carefully choose which of them was preferable for the desired cellular function. Polymerases such as Polß or Polλ, devoted to high fidelity tasks on pre-aligned substrates (BER and NHEJ of highly compatible ends, i.e. those resulting from the repair of staggered nicks) eventually favored the presence of a long nail to help them polymerize correctly, avoiding loss of sequence by dislocation of the template strand. On the other hand, polymerases such as Pol*μ*, designed to use substrates with minimal complementarity, have engineered a long Loop 1 to help them stabilize the connection, even with no complementarity, minimizing the loss of sequence in this step, at the cost of losing the nail. As an additional step in the evolutionary direction taken by each group of polymerases, the rate of single-base substitutions is much lower for Polß and Polλ than for Pol*μ* or *Sc*Pol4 (Bebenek et al., 2005; Roettger et al., 2004), the latter two polymerases having a long Loop 1 and lacking the nail module.

**3. Mt-PolDom: convergent handiness.**

Conventional replicative and lesion bypass DNA polymerases extend off dsDNA substrates, containing both primer and template strands, in a 5' to 3' direction. In contrast, polymerases involved in DSB repair must be capable of binding to and extending off non-canonical DNA polymerase substrates, including 3' over-hanging termini lacking continuous primer and template strands. Recent studies on the bacterial NHEJ polymerases have

**Figure 89. Loop 1 and the nail are mutually exclusive due to steric hindrance.**
Superimposition of the TdT (light pink, 1JMS) and Polλ (light teal, 1XSN)structures showing the DNA substrates in green sticks, and the nail and Loop1 in semi-transarent surface. in pink or red, respectively.



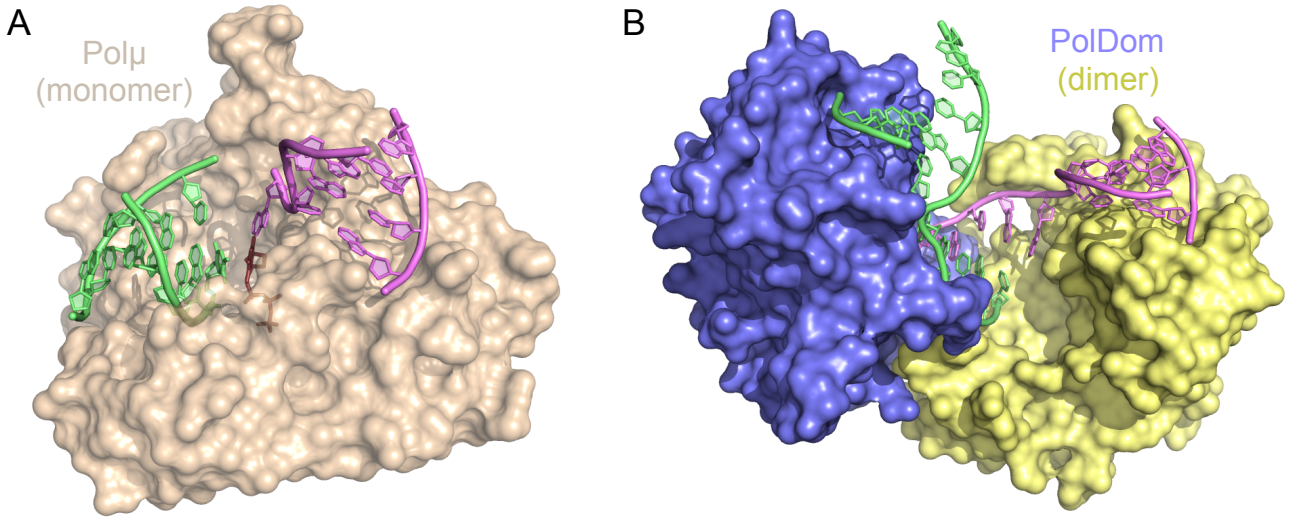



**Figure 90. Different solutions for the NHEJ polymerases: monomers and dimers.**
A) Surface representation of a monomer of Pol*μ* holding two pieces of DNA (green and mauve). B) Surface reresentation of a dimeric arrangement of *Mt*-PolDom (yellow and blue monomers) bridging two DNA ends (green and mauve).

revealed some of the unusual activities associated with these repair enzymes that enable DNA extension under the most extreme conditions. For example, a homodimeric arrangement of the mycobacterial NHEJ polymerases can facilitate the association of two incompatible 3'-protruding DNA ends, *via* microhomology-mediated synapsis, forming a stable end-joining intermediate (Brissett et al., 2007). This synaptic complex reflects an intermediate bridging stage of the NHEJ process, prior to end-processing and ligation. In this way, the polymerase restores the continuity of the dsDNA helix, catalyzing a conventional 5´-3´extension reaction occurring on one DNA end, but templated in trans by a second (synapsed) DNA end. This structure showed an intrinsic difference with the eukaryotic system: working as a dimer versus a monomer, a two-handed versus a one-handed way of fixing broken DNA (Fig. 90). Despite this, and the different origin of the prokaryotic and eukaryotic NHEJ polymerases (AEP family of primases versus X family of DNA polymerases, respectively), we will discuss how these two systems share an unexpected amount of functional and structural features, making it a striking example of convergent evolution.

PolDom is a unique polymerase with a variety of activities on different NHEJ DNA substrates, displaying terminal transferase activity on blunt and ssDNA substrates and templated polymerization: directed *in cis* on gapped and 5'-protruding substrates (Della et al., 2004; Pitcher et al., 2007; Zhu et al., 2006), and *in trans* on 3'-protruding substrates (Brissett et al., 2011; Brissett et al., 2007). The architecture of the bacterial NHEJ polymerases is different to that of the eukaryotic NHEJ polymerases from the X family, although the triad of metal-chelating aspartates is conserved and structurally overimposable (Fig. 91A), a suggestion of the convergent evolution leading to similar catalytic mechanisms. But the convergence does not stop there: in all the activities tested, PolDom shows a marked preference for the insertion of ribonucleotides over deoxynucleotides. This preference, a consequence of the origins of PolDom from the AEP family of primases, reflects a catalytic plasticity that is maintained during evolution on other unrelated NHEJ polymerases such as Pol*μ* (Nick McElhinny and Ramsden, 2003; Ruiz et al., 2003), and now serves a different purpose: to take advantage of the most abundant substrates during a laborious reaction. And, like the eukaryotic NHEJ ligase, LigD can ligate DNA containing ribonucleotides at the 3'-OH terminus (Della et al., 2004; Yakovleva and Shuman, 2006).

Another example of the common characteristics of the prokaryotic and eukaryotic NHEJ polymerases is the presence of a binding pocket for the 5'-P group of the downstream piece of DNA (Fig. 91B). This pocket, which contains residues $Lys^{16}$ and $Lys^{26}$ is missing in AEPs from *Archaea* and *Eukarya* (Fig. 63) is the major determinant for specific binding of PolDom to its substrates and the interaction significantly enhances its activity (Pitcher et al.,

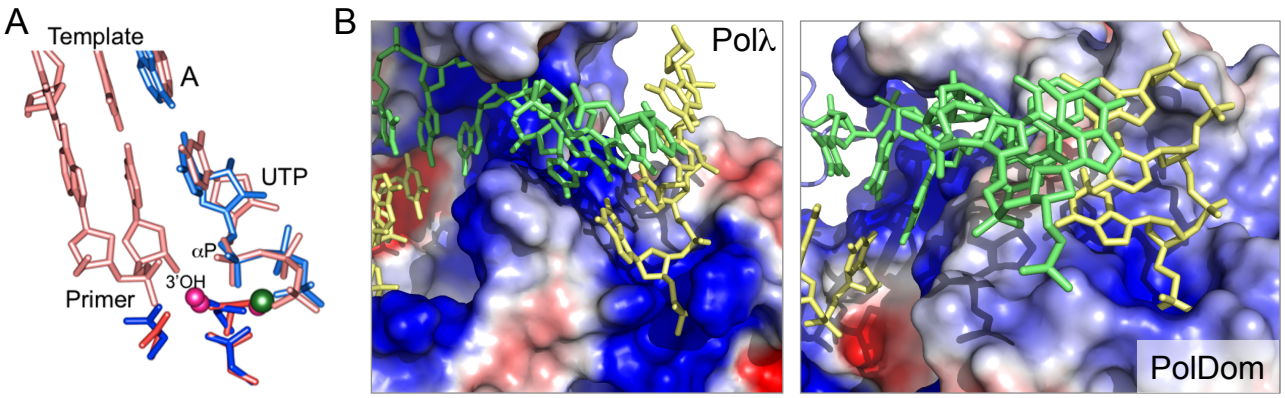



**Figure 91. Similarities among the eukaryotic and prokaryotic NHEJ polymerases.**
A) Superimposition of the ternary complex of Polλ (1XSN) and the pre-ternary complex of Mt-PolDom (3PKY). B) Electrostatic surface of the Polλ and Mt-PolDom 5'-P binding pockets. DNA substrates are shown in green (template strand) and yellow (primer and downstream strands).

2007). While Pol*μ* or Polλ use a specific HhH motif at the 8 kDa domain to bind the phosphate, PolDom lacks this HhH and must therefore utilize a novel structural element to facilitate this interaction.

Although recent studies have provided unique insights into polymerase-mediated orchestration of break synapsis, the order of substrate binding events and mechanism by which these NHEJ polymerases catalyze end-extension is still poorly understood. To address this question, during this PhD thesis and in collaboration with Prof. Doherty lab, we have elucidated the functional meaning of a novel crystal structure of a pre-ternary intermediate of Mt-PolDom bound to DNA, showing that this complex is relevant for specific DSB repair processing events. This catalytically competent complex consists of a PolDom monomer, containing two metal ions and a templated nucleotide (UTP) in its active site, bound to a dsDNA end with a 3' overhang but, significantly, lacking a primer strand. To our knowledge, this structure represents a unique example of a polymerase-DNA complex captured in a pre-ternary intermediate state, relevant for NHEJ.

Is the pre-ternary complex physiologically relevant for prokaryotic NHEJ polymerase extension reactions? Although the pre-ternary complex lacks an incoming primer strand, which provides the attacking nucleophile (3'-OH), a comparison of the positioning of the nucleotide base, phosphate tail, active site ligands and divalent metal ions to those in the active site of a ternary polymerase complex (Polλ) provides compelling evidence that the PolDom pre-ternary complex is catalytically competent (Fig. 91A), even in the absence of the primer strand. The possibility of preforming a pre-ternary complex in solution by incubating the necessary components (PolDom, DNA end, complementary nucleotide and activating metal ions) in the absence of a primer, allowed us to demonstrate its physiological relevance in accelerating NHEJ reactions, probably by providing a "ready to use" primer binding site. By testing the activity of the pre-ternary PolDom complex with different ssDNA primers, we concluded that the minimal primer utilizable by these enzymes is a dinucleotide, as PolDom was not proficient at polymerizing off a single nucleotide "primer". The fact that PolDom cannot efficiently use a single nucleotide as a primer indicates that, although PolDom is evolutionarily related to replicative AEPs, its physiological activity as a primase has effectively been lost and, instead, these polymerases have evolved to have a more restricted capacity to bind short incoming DNA termini, enabling them to perform more specialized roles in NHEJ break repair processes.

The pre-ternary complex mechanism is reminiscent of the initiation reaction carried out by replicative primases, in which the order of events is very similar: firstly, a binary complex between enzyme and DNA is formed, followed by a pre-ternary complex in which the 3' nucleotide is initially stabilized, followed by recruitment of the 5'-nucleotide (the "primer"), to form a ternary complex (Fig. 92). Thus, as a reflection of their common phylogenetic origins, both NHEJ-AEPs and replicative AEPs can catalyze an unorthodox addition of a ribonucleotide in the 3'-5' direction, followed by (in the case of AEPs) conventional (5'=>3') elongation events. The innate ability of

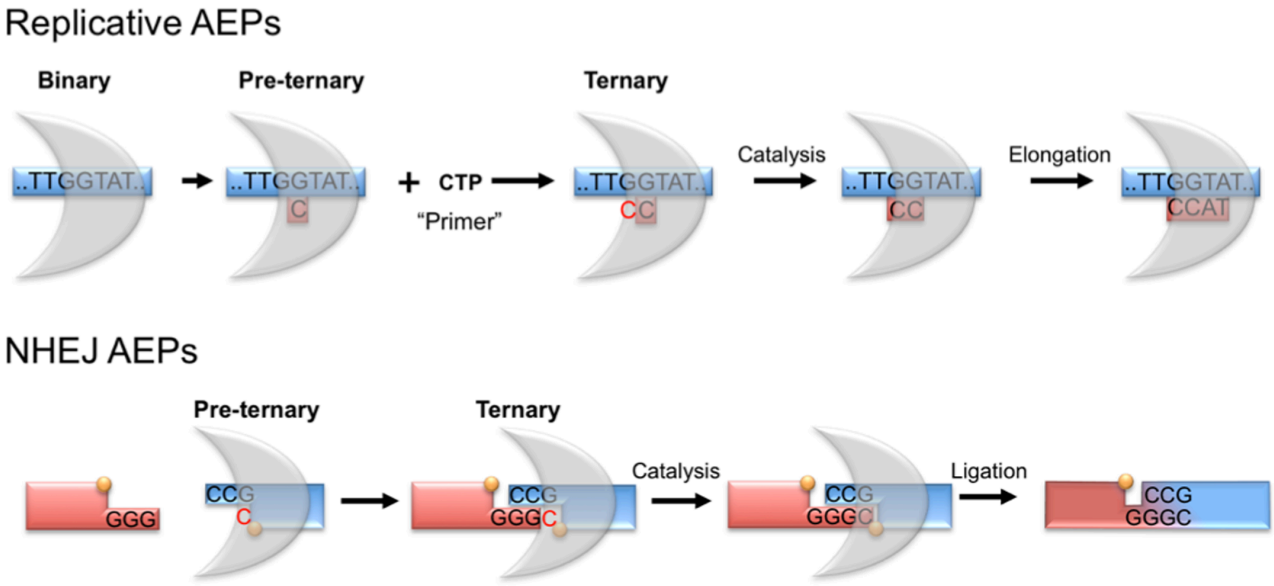



**Figure 92. Comparison of the reaction steps performed by replicative AEPs versus NHEJ AEPs.** (See the main text for details)

AEPs to accept short primers may have influenced evolutionary selection of these enzymes by prokaryotes to become the NHEJ polymerase. Indeed, many bacteria encode additional AEP orthologues whose physiological roles have yet to be determined. Is pre-ternary complex formation also relevant for eukaryotic NHEJ polymerases? It has been demonstrated that human Pol*μ* can catalyze NHEJ extensions on very short and incompatible DNA ends (Davis et al., 2008; Nick McElhinny et al., 2005), a reaction that can take advantage of a limited terminal transferase activity (Andrade et al., 2009), and that can occur with both dNTPs and NTPs. It is likely that formation of a Pol*μ* pre-ternary complex, as that described in chapter 8, triggered by the strong recognition of a 5´-recessive phosphate and a reinforced avidity for the incoming nucleotide (both properties also intrinsic to Pol*μ*) would be beneficial to carry out non-complementary NHEJ of minimally processed ends in eukaryotes, although this remains to be proven.

From a mechanistic point of view, our study of PolDom identified a second conserved loop (loop 2), which plays a prominent role in activation/inactivation of the catalytic centre. The conformation of loop 2 changes significantly, upon the templated-binding of the correct incoming nucleotide, which induces the rotation of the $Arg^{220}$ side-chain (~180°) away from the active site in the pre-ternary complex. Mutation of this invariant residue abolished the extension activity but, significantly, did not alter enzyme binding to other DNA substrates, such as gapped DNA. A comparison of the structures of the PolDom-DNA binary *versus* the pre-ternary complexes reveals the sequential movements that occur in the active site, induced by the binding of both a templating base and an incoming nucleotide. The invariant active site residue $Phe^{64}$, which stacks against the base of the incoming nucleotide in the PolDom-GTP binary complex, now stacks against the base of the templating nucleotide in both PolDom-DNA and pre-ternary complexes, orienting this base and also maintaining (together with $Phe^{63}$) the major kink in the template strand (~105º). In replicative DNA polymerases, aromatic tyrosine residues are commonly employed as a part of a fidelity mechanism that scrutinizes pairing of the correct incoming base with the templating base, thus acting as a molecular gatekeeper to limit the incorporation of an incorrect/mismatched base during elongation (Johnson et al., 2003). We propose that an analogous fidelity mechanism involving the two invariant phenylalanine residues also occurs in the bacterial NHEJ polymerases, but in the absence of the primer strand, thus ensuring that the correctly templated incoming base is bound in the active site prior to the encounter with the incoming end/primer providing the attacking 3'-OH.

This phenylalanine-mediated ($Phe^{64}$) stacking interaction with the templating base in the pre-ternary complex also promotes the movement of the incoming nucleotide (UTP) into the active site and, together with the loss of specific contacts (e.g. $Arg^{246}$, $Lys^{175}$, $Lys^{52}$) promotes the correct

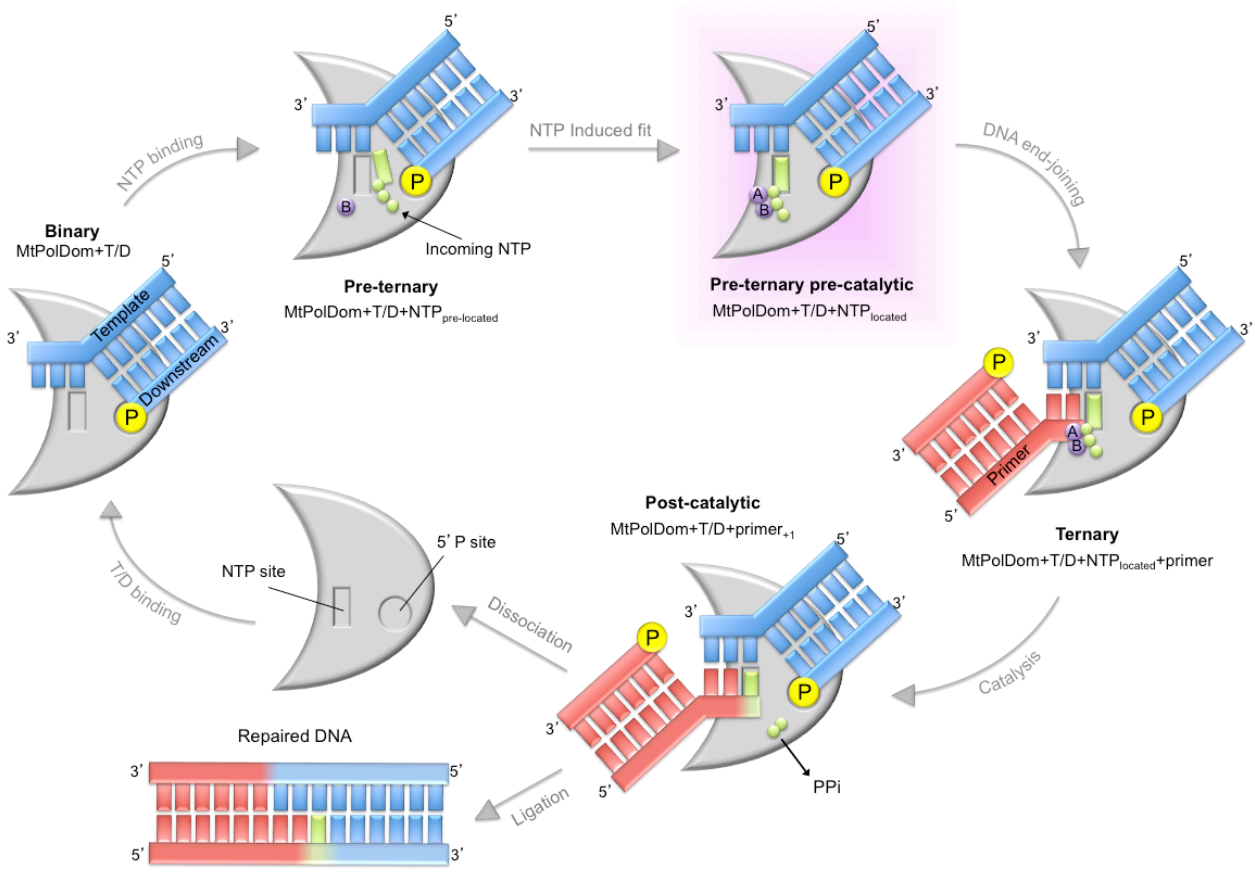



**Figure 93. Catalytic Cycle of a Prokaryotic NHEJ Polymerase**
Initially, a binary complex between the PolDom enzyme (gray crescent) and DNA (T/D; blue) is formed, mainly stabilized via interactions with the 5'-P. Binding of an incoming NTP (green) forms a preternary complex, still incompetent for catalysis as it lacks metal A. Upon template selection and relocation of the complementary NTP and the two metals, A and B, at the correct site (representing a primer-independent NTP induced-fit step) a preternary precatalytic complex is formed. This activated complex, described here, is ready for DNA end joining, allowing the 3'-OH of the incoming primer strand (red) to bind in the active site to form the ternary complex. Further steps of extension, PPi release, dissociation, and ligation (performed by the ligase domain of LigD), complete the DNA repair process.

repositioning of the α-phosphate group of the incoming nucleotide for catalysis. This re-orientated α-phosphate moiety, together with $Asp^{139}$, forms a second metal binding site (A) not present in the binary structure, which is required for the two metal catalytic mechanism common to all DNA polymerases (Brautigam and Steitz, 1998). The binding of the second metal, in turn, promotes breakage of the salt bridge between $Arg^{220}$ and $Asp^{139}$, repositioning this aspartate into a catalytically favorable alignment with the other catalytic aspartates, the α-phosphate group and the two bound metal ions, to form an activated pre-ternary intermediate awaiting the arrival of the nucleophile (3'-OH of the primer strand). The catalytic incompetence of the R220A mutant highlights the importance of the interaction of $Arg^{220}$ with $Asp^{139}$. We propose that the maintenance of this amino acid pairing provides a significant barrier to catalysis until the enzyme becomes optimally bound to DNA, metals, and the correct incoming templated nucleotide. Once these are bound within the active site, a sequence of structural rearrangements promotes the binding of a second metal ion (A). The affinity of $Asp^{139}$ for this second metal promotes the loss of interaction with $Arg^{220}$, leading to expulsion of Loop 2 from the active site, which results in full activation of the catalytic centre. The movement of Loop 2 away from the active site, most likely, promotes this activation step in two ways. The first consequence is that breaking the salt bridge is irreversible, leading to the release of the acidic side-chain of $Asp^{139}$, which is involved in the binding of the second metal (A) within the active site, ensuring that it is optimally poised for catalysis. The second notable consequence, induced by the reorientation of Loop 2, is a significant change in the ridge that surrounds the active site, which most likely allows the incoming 3'-OH of the incoming primer strand to bind in the active site to form the completed ternary

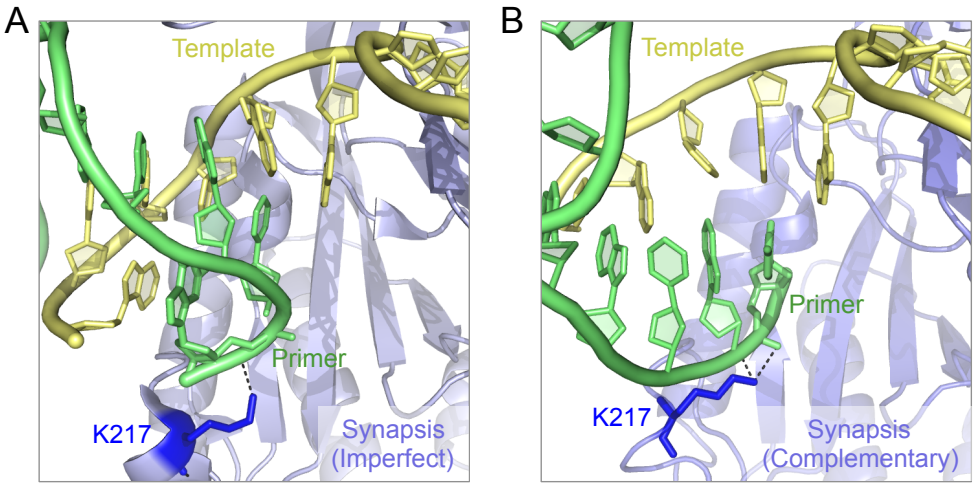



**Figure 94. Lys$^{217}$ as a primer ligand: a versatile and adaptable residue**
A&B) Cartoon representation of the structures of PolDom corresponding to the imperfect synapsis (2R9L, A) and the fully complementary synapsis (B). DNA substrate shown in green (primer) ans yellow (template) sticks, residue Lys$^{217}$ shown in dark blue sticks.

complex (Fig. 62). Further steps of catalysis, PPi release, and ligation would terminate the NHEJ process. A scheme of the different complexes formed during the whole NHEJ cycle are depicted in figure 93. It is remarkable how, despite the different origins of PolDom and Polß, a similar mechanism of prevention of catalysis exists for both of them, as already discussed: an arginine residue contacts one of the catalytic aspartates, keeping it in an unproductive conformation that does not allow catalysis until binding of the nucleotide.

But loop 2 not only has a direct function in the catalytic mechanism through Arg$^{220}$, but also contributes to the stabilization of the two DNA synapsing ends through Lys$^{217}$, which contacts the DNA in the two synaptic complexes obtained to date (Brissett et al., 2007; this PhD thesis). Significantly, the contacts stablished by this residue are similar in the two complexes: both in the imperfect synapsis and in the fully complementary synapsis it is interacting with the primer strand, despite the different orientation of the latter (Fig. 94). Also, mutation of this residue to alanine suggests that the role of this residue differs when the repair reaction can be handled by a monomer (gap substrate) of when a dimer is required (two ends containing 3' overhangs). In the first case, the presence of Lys$^{217}$ is unnecessary or even detrimental for binding to the substrate, while in the second scenario it is completely essential. This evidence and comparison of the structural data indicate that loop 2 configures a landing platform for the binding of the primer, that is remodeled to the "open" conformation during the assembly of the pre-ternary complex to direct the arrival of the primer terminus into the active site. This flexible piece is designed to function specifically during NHEJ, and accordingly, is only present in NHEJ related AEPs (Fig. 63). Other residues help the primer strand to reach its final position, and similar to the function of loop 2 in this regard, are dispensable for gap-filling but fundamental for NHEJ of two DNA ends (this PhD thesis).

We have intensively studied the loops and flexible elements in Pol*μ*, and examined the structure and the mutagenesis studies we have performed on PolDom, reaching the conclusion that both enzymes rely on those movable pieces to perform their most specific activities. As an even more striking example of convergent evolution, PolDom possesses a prominent surface ß-hairpin structure, loop 1, which is specific to NHEJ AEPs (Fig. 63). Conserved residues in loop 1 interact with the 3' protrusion of NHEJ substrates and orient the synapsis of the ends (Brissett et al., 2007; Fig. 70 & Fig. 95A). Mutation of the apical residues of loop 1 to alanine did not affect binding to a primer-containing (gap) substrate, but abolished the ability of PolDom to form a synaptic complex (Brissett et al., 2007) and, consequently, to catalyze trans-directed additions. Loop 1 in Pol*μ* is also specific for binding and activity on NHEJ substrates (Juarez et al., 2006; this PhD thesis), through its function in the stabilization of the synapsis of two DNA ends.

The role of PolDom loop 1 in the stabilization of the template strand is helped by two phenylalanine residues, Phe$^{63}$ and Phe$^{64}$, that maintain the kink in the DNA backbone through stacking interactions with the templating base and the following base, already paired to the 5'-P containing downstream nucleotide (Fig. 95B). According to our experiments, these two amino-acids negotiate the selection of the templating base in the case of existing more than

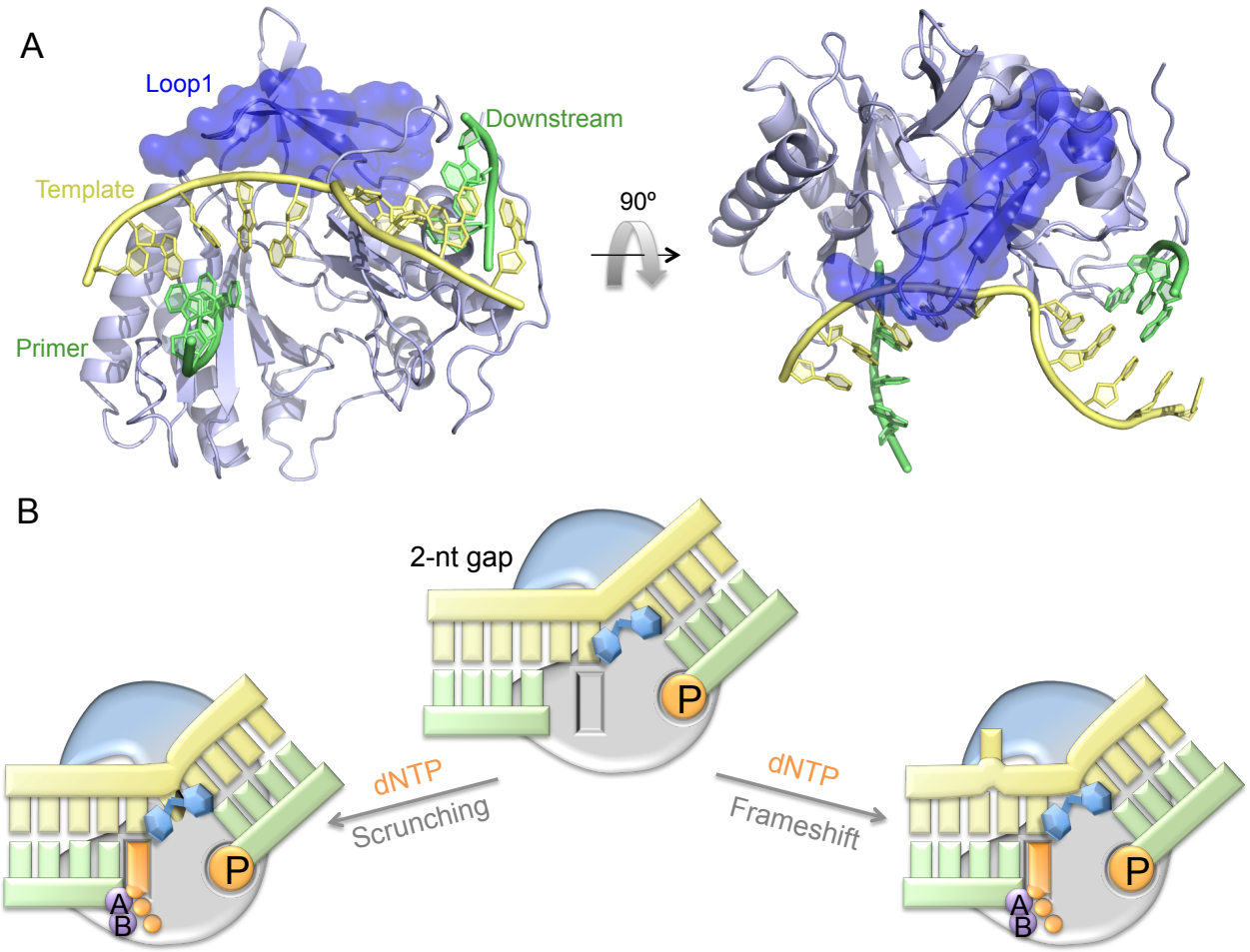



**Figure 95. Selecting the templating base: roles of Loop1 and residues Phe$^{63}$ and Phe$^{64}$.**
A) Cartoon representation of the PolDom-mediated fully complementary synapsis, with the DNA substrate shown in yellow (template strand) and green (primer and downstream strands) and the loop1 shown as a dark blue, semi-transparent surface. B) Cartoons showing the dichotomy that PolDom confronts when dealing with gaps longer than 1 nt: the template strand is either "scrunched", and the gap filled-in correctly (left side), or the template strand is displaced and sequence is lost with the production of frameshifts (right side). The protein is shown as a grey surface with a blue section indicating the approximate position of loop1, incoming nucleotide is colored irange, the two metal ions are shown in purple and the DNA substrate is shown in yellow (template strand) and green (primer and downstream strands). The two phenilalanines are shown as blue hexagons holding the kink in the DNA substrate. See main text for details.

one candidate. It has been shown that in gapped substrates, PolDom has the ability to dislocate and realign the template, extending the primer by inserting nucleotides complementary to templating bases distal to the primer terminus (Pitcher et al., 2007; Yakovleva and Shuman, 2006). This behavior stems from the intrinsic capacity of PolDom to dislocate one or more proximal templating bases, generating base substitutions and frame-shift deletions. The ability to dislocate and accept distorting nucleotides is important to maximize the opportunities to bridge two protruding 3′ ends with limited complementarity. Pol$\mu$ also has a similar template dislocation activity and an ability to realign mismatched ends (Ruiz et al., 2004; Zhang et al., 2001). When these two PolDom phenylalanines (Phe$^{63}$ and Phe$^{64}$) where mutated to alanine, each mutant displayed a different phenotype during adjustment of the templating base in NHEJ reactions: the mutant lacking Phe$^{63}$ was unable to dislocate the first templating base and forced to select the "scrunching" option; Conversely, the mutant lacking Phe$^{64}$ forced the dislocation option, being unable to perform correct scrunching of the second templating base. Thus, the presence of these two residues allows PolDom to choose between one outcome or the other, probably depending on the level of complementarity of the two DNA ends, and to use less microhomology in order to avoid unnecessary loss of sequence, since it is able to correctly polymerize on gaps longer than one nucleotide after bridging. Moreover, the potential to flip-out either the first or the second templating base in this context is of great importance in order to

accommodate mismatches or damaged bases that cannot be used as template during NHEJ reactions. Loop 1 is also implicated in this decision, as supported by our evidence with the triple-alanine mutant: when filling in a 1 nt gap the efficiency was comparable to that of the wild type PolDom, but when confronted with a 2 nt gap in which PolDom incorporated preferentially the nucleotide complementary to the first templating base (scrunching) and showed a low level of incorporation of the second nucleotide (dislocation), the mutant maintained the dislocation levels but had a strongly reduced scrunching ability. Therefore, we propose that loop 1 promotes scrunching of the template strand, allowing PolDom to “count” the templating nucleotides one by one. In this regard, PolDom loop 1 is acting as the nail in Polλ, and enzyme that, unlike Pol*μ*, also has a "counting ability" in a long gap.

## 4. Evolution of the X family of DNA polymerases.

Currently five polymerases have been identified in *E. coli* and at least eight in *S. cerevisiae*, nine in *S. pombe* and fourteen in humans (Bebenek and Kunkel, 2004; Burgers et al., 2001; Hubscher et al., 2002; Pavlov et al., 2006). Based on the primary structure of the catalytic subunits, DNA polymerases have been classified into different families. Eukaryotic organisms have four families: A family (Polγ, Polθ and Polυ), B family (Polα, Polδ, Polε and Polζ) X family (Polß, Polλ, Pol*μ* and TdT) and Y family (Polη, Polι, Polκ and Rev1), whose members were discovered in the last decade (Ohmori et al., 2001), and are involved in replication through DNA lesions. Another significant development was the discovery of Polλ (Garcia-Diaz et al., 2000) and Pol*μ* (Dominguez et al., 2000), which doubled the number of known enzymes of the X family of DNA polymerases, whose members are involved in processes of DNA repair and variability.

### 4.1 Phylogeny of X family polymerases.

The members of this family are present in many organisms, and are found in all monophyletic taxa: Eukarya, Bacteria and Archaea, and even viruses with DNA genome (Oliveros et al., 1997). The high degree of conservation at the structural and amino acid sequence levels between X family members suggests that they originate from a common ancestor.

Unlike what happens in viruses, prokaryotes and yeast, higher eukaryotes have more than one member of the X family. However, there are species in which no member of this family has been described, like the model organisms *Caenorhabditis elegans* and *Drosophila melanogaster* (Burgers et al., 2001), so it becomes a matter of special interest to learn how they have solved the absence of these DNA polymerases in DNA repair processes. Recent data indicate that recombination repair protein 1, the Drosophila homolog of human AP endonuclease 1 (APE1), interacts with DNA polymerase ζ (Takeuchi et al., 2006). It is possible that in protostomes (which include insects and nematodes), APE1-like genes are able to recruit a DNA polymerase other than an X family enzyme to AP sites on DNA. It has been proposed that protostomes evolved from organisms in the coelenterate phylum that lost a Polλ-like gene before other X family DNA polymerases were derived, since it is unlikely that multiple X family genes disappeared as soon as coelenterates appeared (Uchiyama et al., 2009).

Figure 96 shows the phylogenetic relationships between the known members of X family from different organisms. The phylogenetic tree was made using a short and highly conserved segment of the polymerization active site, in order to avoid the presence of accessory domains or small insertions or deletions that may interfere in the analysis. The results suggest that the several subfamilies that can be identified within the X family (Polß, Polλ, Pol*μ* and TdT) have evolved from a common ancestor, perhaps to accommodate different functional requirements. The emergence of more complex organisms seems to promote the specialization of the X family members in order to increase the efficiency of the DNA synthesis processes in which they are involved. The distribution of X family DNA polymerases among different species suggests that the ancestor of the X family DNA polymerase was a Polλ-like gene, which diversified into Polß, Pol*μ* and TdT during evolution. Polλ would have originally been involved in NHEJ to eliminate DNA damage. Subsequently, other X family DNA polymerases were generated in some animals and fungi through gene duplications and they acquired novel roles in DNA metabolism such as in BER and V(D)J recombination. According to very recent results by

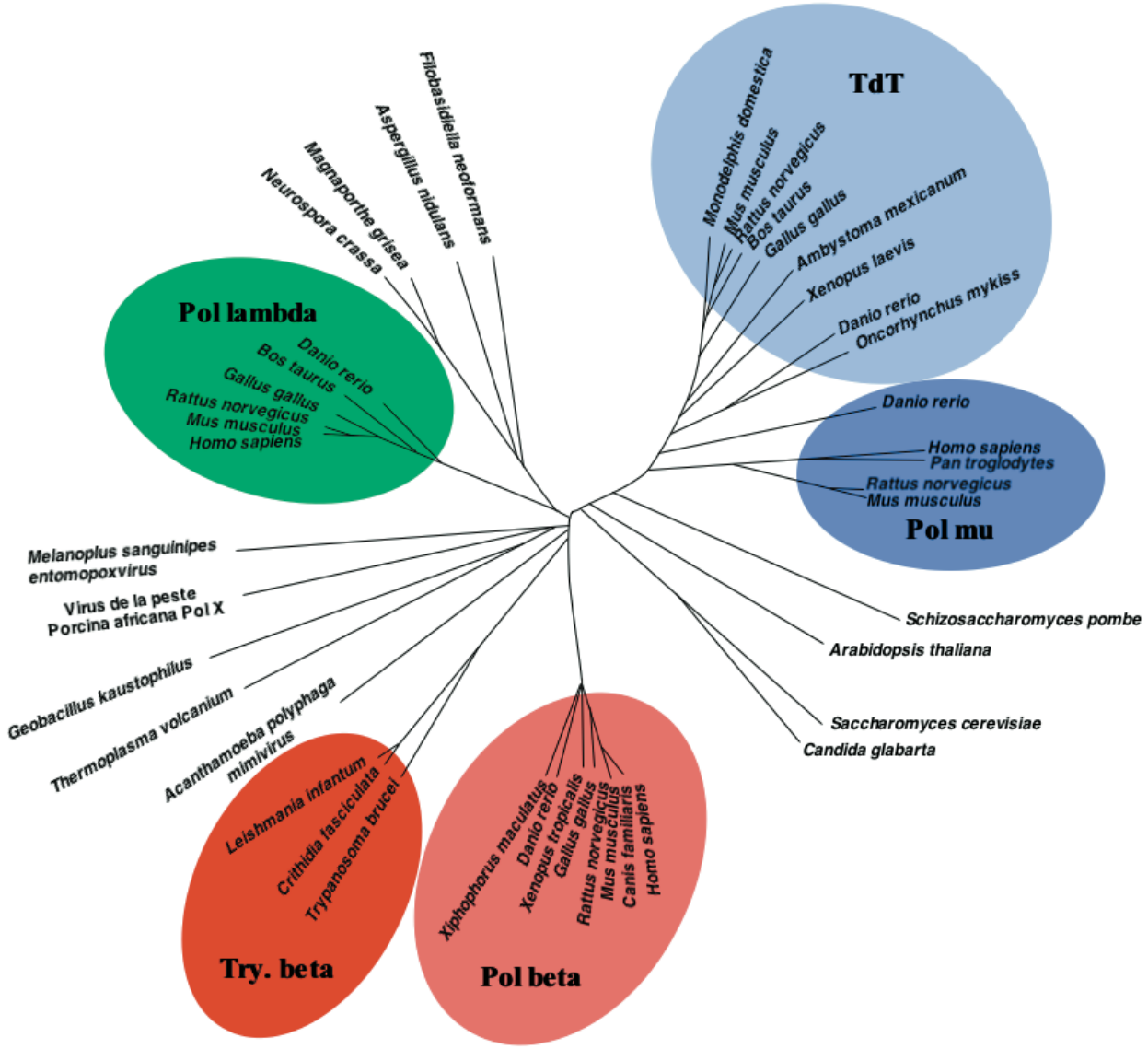



**Figure 96. Evolutionary relationships between family X members.**
An unrooted phylogenetic tree built using a primary sequence alignment of a segment of the catalytic domain. Different enzymes are grouped (shaded areas) into the five main enzyme classes in the family: Polß, Pol, Pol$\mu$, TdT, and trypanosomatid (Try) Polß-like enzymes. Modified from Garcia-Diaz et. al, 2005.

Demogines et al., 2010, these evolutionary forces driving creation of new polymerases are still taking place among primates: codon-based models of gene evolution yielded statistical support for the recurrent positive selection of Polλ, among other four NHEJ genes during primate evolution: XRCC4, NBS1, Artemis, and CtIP. Moreover, analysis of the mutations on the crystal structures available for XRCC4, Nbs1, and Polλ show that residues under positive selection fall exclusively on the surface of these proteins. Studies of positive local evolution on human populations show that, indeed, a single allele of Polλ has previously been reported to be under positive selection in both Asian and Sub-Sahara African populations (Kelley et al., 2009). Also, sliding-window analyses and pairwise comparisons of several strains of *Saccharomyces* indicated that several of the yeast NHEJ genes show evidence of positive selection, including POL4 (Sawyer and Malik, 2006). A first hypothesis explaining the high level of positively selected mutations implies that a certain NHEJ components evolve, compensatory mutations may arise in other NHEJ components to re-optimize protein-protein interactions between the various components. On the other hand, many viruses such as adenovirus and retroviruses like HIV, interact with the proteins of the NHEJ pathway as part of their infectious lifecycle (Bruton et al., 2007; Evans and Hearing, 2005; Jayaram et al., 2008; Kilzer et al., 2003; Li et al., 2001; Lin and

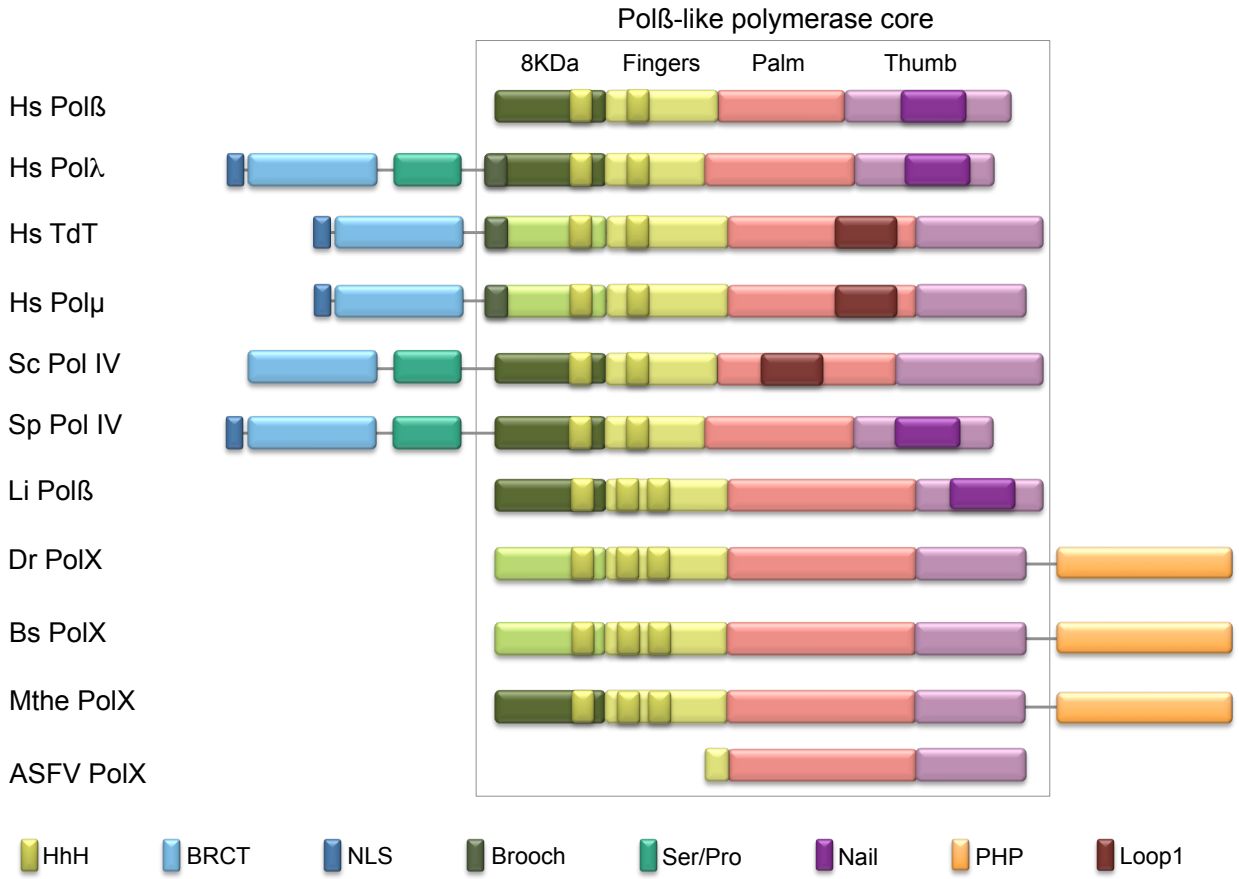



**Figure 97. Modular organization of the family X polymerases.**
Schematic representation of the domains present in family X members from viruses to higher eukaryotes. Regarding the colouring of the 8 kDa domain, dark green represents dRP lyase-containing domains while bright green colour indicates the lack of such activity. *Sc*: *Saccharomyces cerevisiae*; *Sp: Schizosaccharomyces pombe; Li: Leishmania infantum; ASFV: african swine fever virus; Bs: Bacillus subtilis; Mthe: Methanobacterium thermoautotrophicum; Dr: Deinococcus radiodurans.*

Engelman, 2003; Stracker et al., 2002; Weitzman et al., 2004). The Corndog and Omega bacteriophages of mycobacteria have even incorporated the first gene of the bacterial NHEJ pathway, Ku, into their own genome (Pitcher et al., 2006). This viral Ku now evolves under the selective pressures of the virus in order to recruit the bacterial NHEJ ligase, LigD, to circularize phage DNA. Therefore, a second hypothesis would explain the surprisingly rapid evolution of NHEJ genes as an ongoing evolutionary arms race between viruses and these critical genes.

### 4.2 Relationship between structural and functional evolution.

The modular organization of different members of the X family from viruses to eukaryotes indicates the existence of a conserved core Polß type (Fig. 97), the minimal version of it being the PolX from the African swine fever virus (ASFV), which presents only the palm and thumb subdomains of the polymerase domain (Oliveros et al., 1997). The absence of the 8 kDa domain of both ASFV PolX and MSEV (*Melanoplus sanguinipes* entomopoxvirus), may reflect the existence of other proteins encoded by the viral genome to provide the catalytic (dRP lyase) and/or DNA binding properties residing in this domain in most of the DNA polymerases of the X family. Despite the small size of ASFV PolX, it has a second enzymatic activity: the AP-lyase, indicating a possible role in the viral BER pathway (Garcia-Escudero et al., 2003). The evolutionary divergence of the members of the X family has occurred largely by acquisition of additional domains with regulatory properties and/or enzymatic activities. X family members from eubacteria (*Bacillus subtilis*) and Archaea (*Methanobacterium thermoautotrophicum*) have a phosphodiesterase domain (PHP, Fig. 97) fusioned to the Polß core domain, and thus possess polymerase and nuclease activities in the same polypeptide, a great functional benefit to carry out repair processes in the BER pathway.

In eukaryotes there are members of this family from protozoa (*Leishmania infantum*) to mammals. However, there are major differences in the accessory domains that keep a very close relationship with their physiological function. The percentage of similarity at the amino acid sequence level of the Polß core between different members of this family varies from 91% between the Polß enzymes from *Crithidia* and *Leishmania* (LiPolß), and 42% between Pol*μ* and TdT, to 19% identity between LiPolß and TdT (Taladriz et al., 2001). LiPolß shows a 31% of amino-acid identity with mammalian Polß, close to the 32% between Polλ and Polß. Interestingly, both Polß enzymes from *Crithidia* and *Leishmania* present inserts within the core that allow protein-protein and protein-DNA interactions. Contrary to mammals, yeast have a single DNA polymerase from the X family, Pol4. Both Pol4 from *S. cerevisiae* and *S. pombe* possess two additional domains at their N-terminus: a BRCT domain followed by a regulatory Ser/Pro domain (Fig. 97). In addition, both Pol4 have a dRP-lyase activity associated with the 8 kDa domain suggesting a role for yeast Pol4 in repair processes such as BER (Bebenek et al., 2005; Gonzalez-Barrera et al., 2005). Although both Pol4 enzymes share a common structural organization, their differ in terms of sequence similarity with their human counterparts. While *Sc*Pol4 is more similar in the composition of the basic Polß structure to Polλ, sharing a 25% of amino-acid identity (Bebenek et al., 2005), *Sp*Pol4 is closer to Pol*μ* (27% amino-acid identity) than to Polλ (24% amino-acid identity). Based on sequence similarity one can speculate that in yeast *Sp*Pol4 is the orthologue of human Pol*μ* while *Sc*Pol4 could be the orthologue of human Polλ.

The presence of BRCT domains in Pol4, Polλ, Pol*μ* and TdT, relates to the role that this domain plays in processes such as V(D)J recombination and NHEJ repair. The BRCT domain of Pol4 mediates the interaction of the polymerase with factors involved in the NHEJ pathway during repair of double strand break (Tseng and Tomkinson, 2002; Tseng and Tomkinson, 2004). Similarly, the BRCT domain of Polλ, Pol*μ* and TdT in higher eukaryotes, allows these proteins to participate in both NHEJ repair and V(D)J recombination. It is possible that subtle differences in the amino acid sequence of the BRCT domain of each polymerase have great importance in regulating the access of each DNA polymerase to a specific substrate or protein of the route.

Finally, the eukaryotic Polß (initially thought to be exclusive of mammals) has lost some accessory domains during evolution, a crucial step for its specialization as a housekeeping DNA repair polymerase that protects against the large amount of oxidative damage present as a result of aerobic metabolism. The conservation of the polymerase domain and the 8 kDa domain (Fig. 97), where the dRP-lyase activity lies, is central for participation in the BER pathway.

The similarity between yeast Pol4 and Polλ, which share the same additional domains (Fig. 97), together with the extraordinary evolutionary conservation of the versions of Polλ present in various higher eukaryotes and in plants (*Arabidopsis thaliana, Wisteria max, Oryza sativa*) suggests that this is the X family member closest to the common ancestor from which all members of the family derived. This could account for the multiple functions of Polλ, since the common ancestor necessary carried out various processes of DNA synthesis. In this sense, the presence of the Ser/Pro domain is of special relevance, as it could regulate the participation of Polλ in different processes, such as repair by BER, NHEJ and V(D)J recombination.

Members of the human X family of DNA polymerases have specialized in different processes of DNA synthesis associated with repair. Such processes are basically three: 1) base excision repair (BER), carried out mainly by Polß, although Polλ seems to have a role in specific situations; 2) non-homologous end joining (NHEJ), in which, according to the type of substrate generated, Polλ or Pol*μ* could be involved; 3) V(D)J recombination, involving Polλ, Pol*μ* and TdT, with different roles. Subtle differences in the biochemical properties of X family members seem to be crucial for performing one role and not other. Therefore, the members of this family have diversified to be able to carry out non-redundant tasks, achieving a high degree of specialization that has resulted in a high degree of efficiency of each polymerase on its specific function.

Polλ, as the member of the family more closely related to the common ancestor, bears many of the specific modifications needed to perform a high number of functions: it has a BRCT domain needed for interactions with the NHEJ components, and it

harbors an 8 kDa domain that acts both as the main DNA binding domain through the 5'-P pocket and the container of the dRP-lyase activity needed for an efficient performance during BER. Moreover, it contains a long nail that helps the polymerase to deal with misaligned substrates and might allow scrunching to occur, it has a brooch (WxCxQ motif) that maintains the Polß-like core in a closed conformation throughout the catalytic cycle possibly helping to correctly orient discontinuous NHEJ substrates, and finally it has a mid-length Loop 1 that may have a similar role to that proposed for Pol*μ* Loop 1 during NHEJ, but with the limitation of needing some degree of complementarity between the two DNA ends, probably due to the position occupied by this loop in Polλ at the -2 to -4 positions of the template strand.

As the youngest member of the family, Polß is the polymerase that has lost the majority of these features, in order to be focused in enhancing the efficiency of just one reaction: the filling-in of short gaps during BER. For that, it has strengthened the interactions with the DNA substrate through the 5'-P binding pocket, being the most positively charged in this region of the four human enzymes, and it has maintained the dRP-lyase activity and gained an AP-lyase activity, precious for its dedicated job as a BER polymerase. It also maintains a long nail that helps locating the DNA substrate on its final catalytic position, and probably helps to "count" the templating nucleotides when filling-in a long gap. It also has the capacity of changing from an "open" to a "closed conformation" since it has lost the brooch at the N-terminal portion of the core, and thus the space between the 8 kDa domain and the thumb subdomain can be expanded to accommodate the yet-to-be-copied templating nucleotides more easily. On the other hand, the loss of this "closing" motif probably meant that its role as a NHEJ polymerase was greatly impaired, together with the complete loss of the Loop 1, that is now merely a turn connecting two ß-strands. The disappearance of this flexible structure probably also led to an improvement of the polymerization on template-containing substrates such as the ones produced during BER. Finally, Congruently, Polß lost the BRCT domain so it does not get recruited to DNA DSBs where it cannot act, and has in turn gained a new set of protein-protein interactions with other BER factors as XRCC1 through specific residues on the surface of its catalytic domain that are required for an efficient repair (Dianova et al., 2004; Gryk et al., 2002; Marintchev et al., 2000). The Ser/Pro domain located between the BRCT and the catalytic domains in Polλ is also missing in Polß, and this, together with the total absence of CDK phosphorylation sites, unique in the human X family, indicate the lack of a cell-cycle dependent regulation that correlates with its function as a housekeeping gene. Whereas short-patch BER in mammalian cells plays an important role in the maintenance of genome stability (Fortini et al., 1998; Sobol et al., 1996; Sobol and Wilson, 2001), it is unlikely that a similar repair pathway is present in many phylogenetically divergent organisms. Plants do not contain a homolog of DNA ligase III, which is required for mammalian short-patch BER, or a Polß homolog (Kimura and Sakaguchi, 2006). Additionally, the plant XRCC1 protein lacks the Polß binding domain (N-terminal domain; {Uchiyama et al., 2008). In contrast, all enzymes needed for long-patch BER are encoded in the genomes of *A. thaliana* and *O. sativa*, suggesting that plants utilize the long-patch BER pathway (Kimura and Sakaguchi, 2006). Similarly, no protostomic organism possesses the short-patch BER system (Radford et al., 2007; Takeuchi et al., 2006), and a short-patch BER-like pathway is present in yeast but it differs from the mammalian pathway (Alseth et al., 2005). From the data described above, we hypothesize that short-patch BER is an advanced repair pathway present only in mammals (Fig. 98). Polß, the primary DNA polymerase of this pathway, is highly expressed in brain tissue (Hirose et al., 1989), and would be required mainly to minimize the accumulation of DNA damage in neuronal cells (Sugo et al., 2000) that suffer from a high level of oxidative lesions (Nakamura and Swenberg, 1999; Wilson and McNeill, 2007).

Finally, the last two members of the human family are Pol*μ* and TdT, the most closely related of the four, with a 42% identity at the level of the aminoacid sequence. Although the branch of the phylogenetic tree of the X family that contains these two enzymes appeared much sooner than Polß, the strict template-independent activity of TdT appears to be a recent evolutionary event that coincides with the development of V(D)J recombination in mammals (Fig. 98). TdT shares the common Polß-like core with 8 kDa, fingers, palm and thumb and also

possess he C-terminal BRCT domain that allows recruitment by the Ku proteins to the site of the break. But there are some differences: even though TdT still conserves a positively charged pocket to bind a downstream 5'-P, it contains the lowest amount of positive charges of all the members of the family, and, equal to what happens in Pol*μ*, it has lost the residues essential for the dRP-lyase activity. This first modification, together with the tightly regulated expression of TdT confined to primary lymphoid tissues including thymus and bone marrow (Bollum, 1979; Coleman et al., 1974; Kunkel et al., 1986), already indicates that TdT, even though devoted to work at DSBs, is not able to deal with damaged nucleotides and the break points must be "clean", as they are in the case of programmed breaks such as those occurring during the development of the immune response. TdT has been in fact engineered through evolution to "misbehave" and break almost every rule that can apply to a conventional DNA polymerase: it incorporates nucleotides in a template independent manner, using only single stranded DNA (Bollum, 1959; Bollum, 1964) or dsDNA with a 3'-overhang longer than four nucleotides (Gozalbo-Lopez et al., 2009). This strict preference for the DNA substrate is dictated by its long Loop 1, of about the same length as the one present in Pol*μ*, but immobilized by several interactions not present in Pol*μ*, such as the ones stablished between Loop 1 and the small thumb loop (this PhD thesis). The position of Loop 1 in the crystal structure completely superimposes with the template strand from the Pol*μ* ternary complex, thus explaining why the length of the single stranded portion needs to be of at least 4 nucleotides for an efficient reaction to take place. This protein piece helps locate the nucleotide in place, and probably is to be blamed for the different order of substrate binding displayed by TdT in contrast with other polymerases: efficient

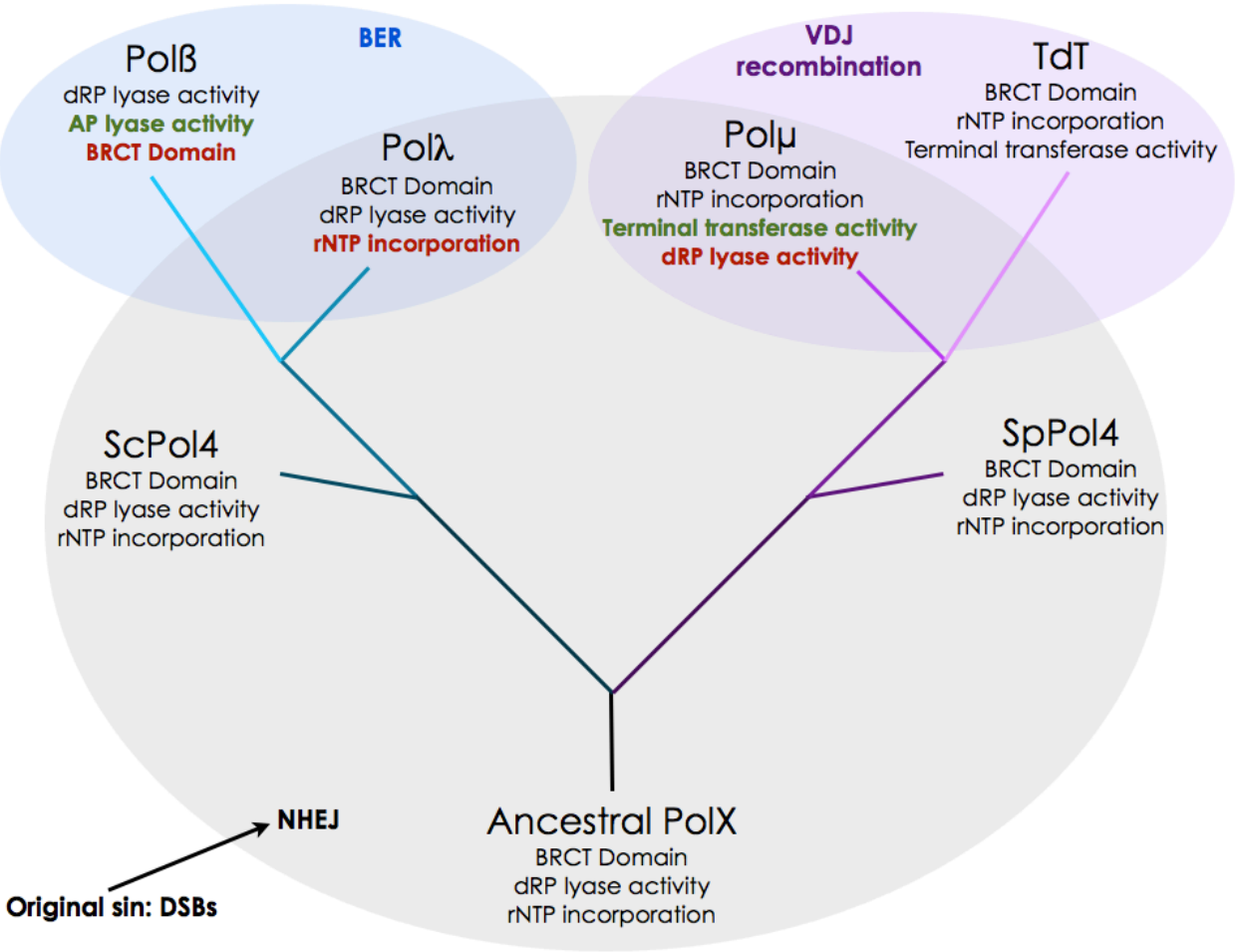



**Figure 98. Evolution of family X polymerases.**
Red colour indicates the loss of an activity or feature, green colour indicates the gain of an activity or feature. See text for details.

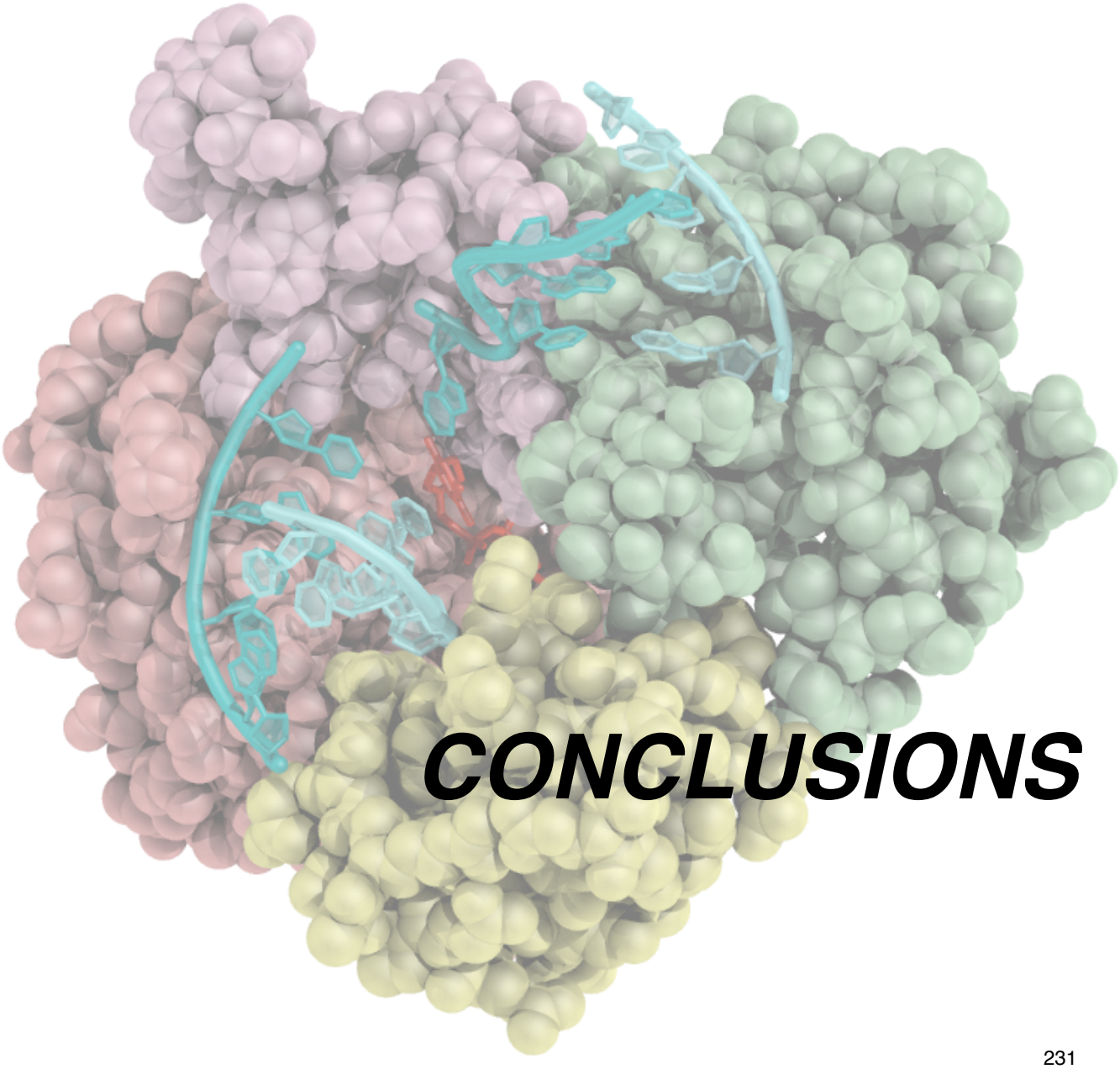
polymerization for a template-dependent polymerase would be optimal through the strictly ordered binding of DNA substrate prior to dNTP, as the converse order of dNTP binding prior to DNA would be error-prone, being correct only once out of four opportunities. Indeed, numerous steady-state and pre-steady state studies have validated that all template-dependent polymerases obey this mechanism (Benkovic and Cameron, 1995). The order by which TdT binds DNA and dNTP is indeed random as determined through a series of initial velocity studies (Deibel and Coleman, 1980): TdT forms the catalytic competent ternary complex *via* binding of dNTP prior to DNA or vice versa. This scenario is similar to that observed for the *Mycobacterium* NHEJ polymerase, in which a pre-ternary complex can be formed with the nucleotide being present in the absence of a primer strand (Brissett et al., 2011). This situation could apply also to Pol*μ*, as it would be beneficial for the efficiency of DSB repair, and could have been maintained in TdT since the ability to randomly bind substrates could play a physiological role in generating random nucleotide additions during recombination. Another feature that is present in Pol*μ* and has been maintained in TdT evolution is the ability to incorporate ribonucleotides. This loss of the "steric gate" probably appeared in Pol*μ* as a collateral effect of the need for a spacious active site able to accommodate misalignments during the search for microhomology, and has been positively selected due to the optimal characteristics of the ribonucleotides as the most abundant substrates, but also due to the "length control" mechanism that the incorporation of ribonucleotides implies during un-templated addition of nucleotides: for both Pol*μ* and TdT, further elongation of a ribonucleotide-containing primer occurs at a slower rate and the addition of more than two ribonucleotides does inhibit activity (Nick McElhinny and Ramsden, 2003; Roychoudhury, 1972; Ruiz et al., 2003).

Despite all the similarities between Pol*μ* and TdT, such as the loss of the dRP-lyase activity, the ability to incorporate ribonucleotides and the presence of Loop 1, as well as the absence of a long nail, Pol*μ* has remained preferentially a template-directed polymerase. In the first place, being a more ancient product of evolution than TdT means that its function had to be a more general one: Pol*μ* was devoted mainly to its function in the NHEJ repair pathway. The differential expression patterns of TdT and Pol*μ* also speak in favor of this hypothesis: even though Pol*μ* is strongly expressed in lymphoid tissues in humans, in contrast to TdT, a basal expression of Pol*μ* is observed in a wide range of tissues, more specifically in the brain (Dominguez et al., 2000), that suffers a high level of oxidative damage. Also, the structural features of Pol*μ* support its role as a template-directed NHEJ polymerase: a flexible Loop 1, contacted but not constrained by several other modules in the protein (the thumb loop, the arginine helix), helps stabilize gaps in the template strand without blocking the use of the templating base. Also, a specific arginine residue ($Arg^{387}$), present only in Pol*μ*, acts as a "brake" during the terminal transferase catalytic cycle, limiting the number of untemplated additions and keeping the polymerase in a "stand-by" mode for a longer time, awaiting the arrival of the templating base.

Taking advantage of the Dr. Jekyll & Mr. Hyde duality of Pol*μ* as a template-directed and also template-free polymerase, its appearance in the phylogenetic tree of the X family probably was the starter's pistol shot to the process of generating variability during development of the adaptive immune system response, without losing the DNA repair function. In fact, it has been demonstrated that Pol*μ* still participates in the $DJ_H$ rearrangements in mice embryos, where TdT is still not expressed (Gozalbo-Lopez et al., 2009), and based on its DNA-dependent polymerization ability, which TdT lacks, Pol*μ* also fills in small sequence gaps at the coding ends and contributes to the ligation of highly processed ends, frequently found in the embryo, by pairing to internal microhomology sites. Also, Pol*μ* is involved in V(D)J recombination at immunoglobulin k light-chain loci, after synthesis of the N-regions (Bertocci et al., 2003). The lack of Pol*μ* leads to alterations that induce a profound defect in the peripheral B cell compartment which results in an average 40% reduction in the splenic B cell fraction in Pol*μ* knock-out mice. Pol*μ* appears, therefore, as a key element contributing to the relative homogeneity in size of light chain CDR3 and taking part in Ig gene rearrangement at a stage where TdT is no longer expressed (Bertocci et al., 2003). Pol*μ* has also been shown to be up-regulated in germinal centers after immunization, and although it is not a critical partner, Pol*μ* modulates the *in vivo* somatic hypermutation (SHM) process (Lucas et al., 2005).

The role of Pol*μ* in this process was proposed some time ago (Ruiz et al., 2001), and further supported by studies of Pol*μ* overexpression in a Burkitt's lymphoma cell line (with constitutive SHM), in which the SHM rate was increased (Ruiz et al., 2004).

In recent years, structural genomics has given rise to a vast array of knowledge, that nonetheless needs to be interpreted correctly as a range of still snapshots of a movie that, if seen, would show the highly complex and ever-moving machines that polymerases are. Helped by the biochemistry, and placed in context by the *in vivo* data, this structural approach has been used here to better understand the unique properties of each of the human DNA polymerases of the X family, and also of their bacterial counterparts. Thorough analysis of these structures has driven this PhD thesis and has provided us with a deeper understanding of the unique activities and specificities attributed to each polymerase.



# CONCLUSIONS

1. The analysis of Pol*μ* DNA binding properties through EMSA and footprinting assays emphasizes the importance of the presence of a phosphate group at the 5'-terminus of the downstream strand, specially when dealing with 3'-protruding NHEJ substrates. We could observe the formation of both the binary and the ternary complex in our assays, demonstrating that, upon binding of the dNTP, a conformational change is induced and a tighter binding to the template/primer area is achieved. We have discovered a new function for the BRCT domain of Pol*μ*, as a DNA-binding domain. The BRCT adds strength to the association of the enzyme with the downstream part of the substrate, as assessed by footprinting assays. Accordingly, NHEJ activity of a deletion mutant lacking this domain was greatly diminished.

2. Regarding NHEJ DNA substrates, and depending on the sequence context, Pol*μ* can efficiently and faithfully insert trans-directed nucleotides in the absence of other NHEJ factors. The presence of a 5'-P group on the downstream strand is critical for Pol*μ* binding and precise joining of incompatible ends. In this context, Pol*μ* is able to accept distortions in the connection, such as mismatches and flipped-out nucleotides without diminishing the efficiency or fidelity of the outcome. Pol*μ* can also use damaged bases such as 8oxoG as template or in the connection of two DNA ends.

3. The previously uncharacterized "brooch" motif (YACQR in Pol*μ* and TdT), potentially implicated in the maintenance of a closed conformation of the Polß-like core of the polymerase, is of specific importance during NHEJ of non-compatible ends: single mutations of the conserved residues completely abolished the ability of Pol*μ* to perform end-joining reactions.

4. We have identified the conditions that lead to a gain in efficiency without a cost in terms of fidelity by using ribonucleotides instead of deoxynucleotides and different activating metal ion combinations. $Mn^{2+}$ ions, which provide a tighter binding of the incoming nucleotide and are known to be optimal activators of Pol*μ* terminal transferase, do benefit NHEJ by improving the efficiency of nucleotide insertion. We have proved for the first time that the accuracy of the enzyme is maintained when $Mn^{2+}$ ions are used at the appropriate physiological concentration. We propose that Pol*μ* could take advantage of a versatile active site, valid for selecting alternative activating metal ions and nucleotides as substrates. This versatility would allow "ad hoc" selection of the most appropriate nucleotide/metal combination for individual NHEJ events, thus widening the spectrum of available solutions to position a templating base in proper register for connection.

5. We have proven that the terminal transferase activity of Pol*μ* is specifically required during non-homologous end joining to create or increase complementarity of DNA ends. By site-directed mutagenesis in human Pol*μ*, we have identified a DNA and dNTP dual ligand, $His^{329}$, necessary for both activities, and a specific DNA ligand residue, $Arg^{387}$, that is responsible for Pol*μ* limited terminal transferase activity compared to that of TdT. This arginine has a regulatory role in the translocation of the primer from a non-productive E:DNA complex to a productive E:DNA:dNTP complex in the absence of a templating base, the rate limiting step during template-independent synthesis. The presence of a second 3'-protrusion, containing a 5'-P group and providing a templating base, facilitates this special "induced-fit translocation step". This mechanism thus regulates the balance between accuracy and necessary efficiency, providing some variability during NHEJ.

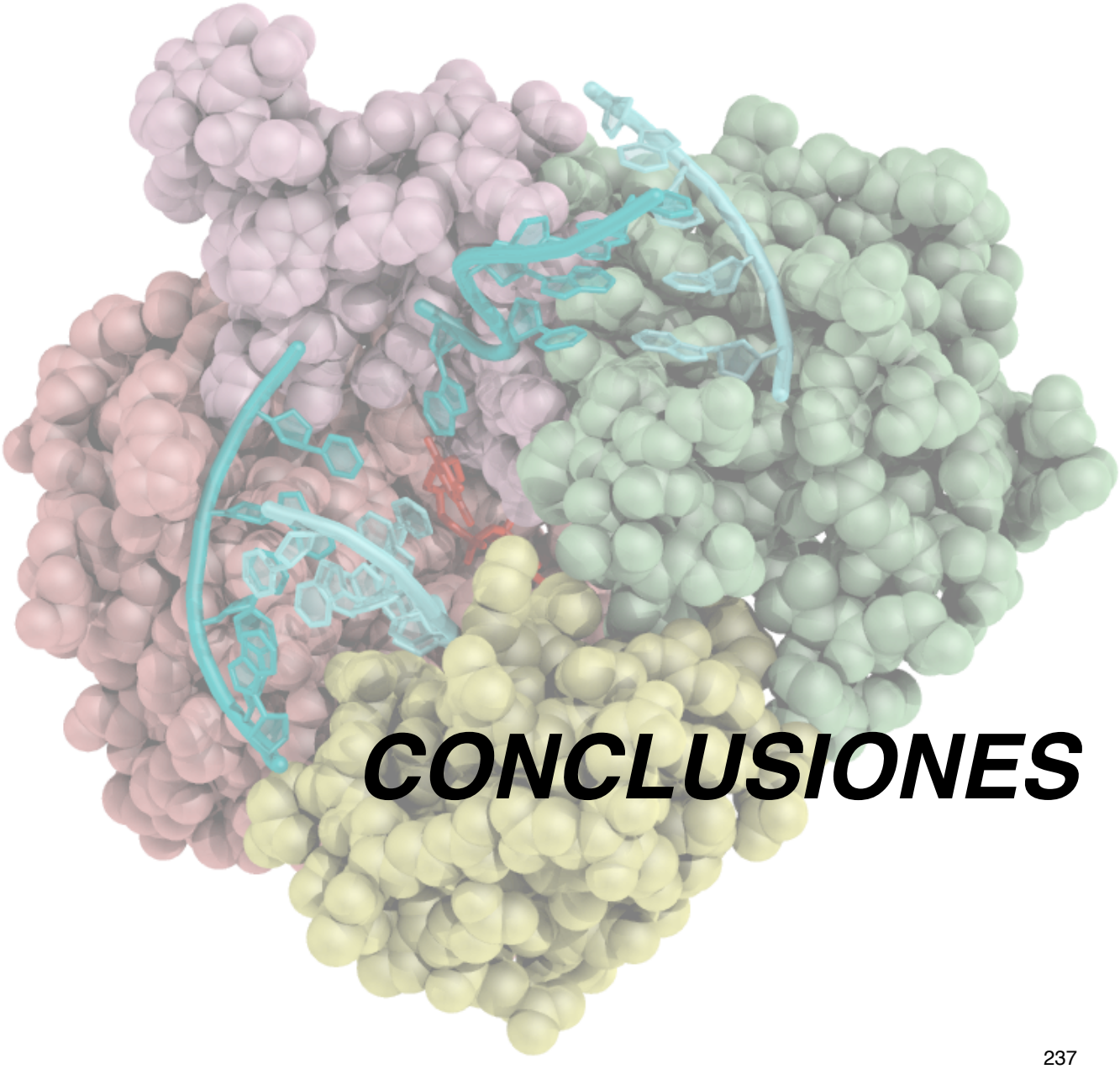
6. Loop 1 is a flexible piece of Pol*μ* that has a critical role during terminal transferase and end-joining activities since it acts as a pseudo-template when the template strand is discontinuous or unavailable, while diffusing away when it is present to avoid steric clashes. Our site-directed mutagenesis results show that the previously uncharacterized thumb “mini-loop” (NSH motif) and the positively charged helix N greatly contribute to the correct positioning of the template strand or Loop 1 in each case. Other point mutations in the “hinges” of Loop 1, such as residues $Phe^{385}$ or $Phe^{389}$, corroborate the flexibility requirements of this motif.

7. We have identified two positively charged regions in Pol*μ*, devoted specifically to the binding of NHEJ substrates as separate entities: the primer binding zone, containing residues $Lys^{249}$, $Arg^{253}$ and $Arg^{416}$, and the template binding zone, formed mainly by residues $Arg^{442}$, $Arg^{445}$, $Arg^{449}$ of the helix N. Single mutation of any of these residues, while unimportant in most of the cases for gap-filling, completely abolished end-joining activity.

8. Concerning the possible post-translational modifications involved in Pol*μ* regulation, we have identified the residues that could be phosphorylated *in vivo*. Our evidence suggests that themain modification event affects mainly the functionality of Loop 1, through phosphorylation of $Ser^{372}$. Therefore, the most distinctive activities of Pol*μ* (i.e. terminal transferase, end joining of non-compatible ends) would be turned off at specific cell-cycle phases (late S and G2), in which Pol*μ* promiscuity might be harmful to the cell.

9. The *Mycobacterium tuberculosis* NHEJ pathway includes an AEP-related enzyme (PolDom) as the main polymerase involved. We have examined a new complex containing one PolDom monomer, a template/downstream DNA substrate, incoming NTP and activating metal ions, and our results show that even in the absence of the primer strand this complex is catalytically competent and beneficial for the subsequent end-joining reaction. This pre-ternary complex also showed that a highly conserved residue, $Arg^{220}$, located in the NHEJ-related enzymes specific loop2, is the main switcher to turn off/on the active site through an inactivating contact with one of the catalytic aspartates.

10. The resolution of a second PolDom structure showed for the first time a complementary synapsis of two DNA ends, maintained by a PolDom dimer. Analysis of the 3D data and our biochemical results indicate that loop2 is not only involved in catalysis, but also in the correct positioning of the incoming primer, through $Lys^{217}$. Moreover, correct selection of the templating base during NHEJ is shown to rely on two phenylalanines, that supply PolDom with the ability either to correctly copy the template or to generate frameshifts, thus providing the versatility required to perform an end-joining reaction.



# CONCLUSIONES

1. El análisis de las propiedades de unión al DNA de Pol*μ* a través de EMSA y ensayos de *footprint* hace hincapié en la importancia de la presencia de un grupo fosfato en el extremo 5' de la cadena *downstream*, especialmente cuando se trata de sustratos de NHEJ 3'-protuberantes. Hemos podido observar la formación tanto del complejo binario como ternario en nuestros ensayos, lo que demuestra que, tras la unión del dNTP, se produce un cambio conformacional y se induce una unión más estrecha con el área *template*/*primer*. Hemos descubierto una nueva función del dominio BRCT de Pol*μ*, como un dominio de unión al DNA. El BRCT añade fuerza a la asociación de la enzima con la parte posterior del sustrato, según la evaluación de los ensayos de la huella. En consecuencia, la actividad NHEJ de un mutante de deleción que carecen de este dominio ha sido disminuido considerablemente.

2. En cuanto a los sustratos de DNA de NHEJ, y dependiendo del contexto de secuencia, Pol*μ* puede, de manera eficiente y fiel, insertar nucleótidos dirigidos en *trans* en la ausencia de otros factores de NHEJ. La presencia de un grupo 5'-P en la cadena *downstream* es fundamental para una unión precisa de Pol*μ* a extremos incompatibles de DNA. En este contexto, Pol*μ* es capaz de aceptar distorsiones en la conexión, tales como nucleótidos extrahélice o apareamientos erróneos, sin disminuir la eficiencia o la fidelidad del resultado. Pol*μ* también puede utilizar bases dañadas como 8oxoG como molde o en la conexión de los dos extremos de DNA.

3. El motivo "broche" (YACQR en Pol*μ* y TDT), no caracterizado previamente, y que puede estar implicado en el mantenimiento de una conformación cerrada del core polimerasa de Pol*μ*, es de especial importancia durante NHEJ de extremos no compatibles: mutaciones individuales de los residuos conservados abolieron completamente la capacidad de Pol*μ* de llevar a cabo reacciones de reparación.

4. Hemos identificado las condiciones que conducen a un aumento en la eficiencia sin un coste en términos de fidelidad mediante el uso de ribonucleótidos en lugar de deoxinucleótidos y diferentes combinaciones de iones metálicos. Los iones $Mn^{2+}$, que proporcionan una unión más estrecha del nucleótido y se sabe que son activadores óptimos de la actividad transferasa terminal de Pol*μ*, benefician el NHEJ mediante la mejora de la eficiencia de la inserción de nucleótidos. Hemos demostrado por primera vez que la fidelidad del enzima se mantiene cuando los iones $Mn^{2+}$ se utilizan en la concentración fisiológica adecuada. Proponemos que Pol*μ* podría tomar ventaja de un sitio activo versátil, válido para la selección de iones metálicos y de nucleótidos alternativos como sustratos. Esta versatilidad le permite la selección "ad hoc" de la combinación más adecuada de nucleótido y metal para los distintos eventos de NHEJ, ampliando así el espectro de soluciones disponibles para el correcto posicionamiento de la base molde en registro para la conexión.

5. Hemos demostrado que la actividad transferasa terminal de Pol*μ* se requiere específicamente durante el NHEJ para crear o aumentar la complementariedad de los extremos de DNA. Por mutagénesis dirigida en Pol*μ*, hemos identificado un ligando dual del DNA y del nucleótido, $His^{329}$, necesario para ambas actividades, y un ligando específicos del DNA, $Arg^{387}$, que es responsable de la actividad de la limitada actividad transferasa terminal de Pol*μ* en comparación con la de TdT. Esta arginina tiene un papel regulador en la translocación del *primer* de un complejo E:DNA no productivo a un complejo E:DNA:dNTP productivo en la ausencia de una base molde, el paso limitante en la síntesis independiente de molde. La presencia de un segundo extremo 3'-protuberante, que contenga un grupo 5'-P y proporcione la base molde, facilita este "ajuste inducido del paso de translocación". Este mecanismo regula el equilibrio entre la fidelidad y la necesaria eficiencia, proporcionando una cierta variabilidad durante el NHEJ.

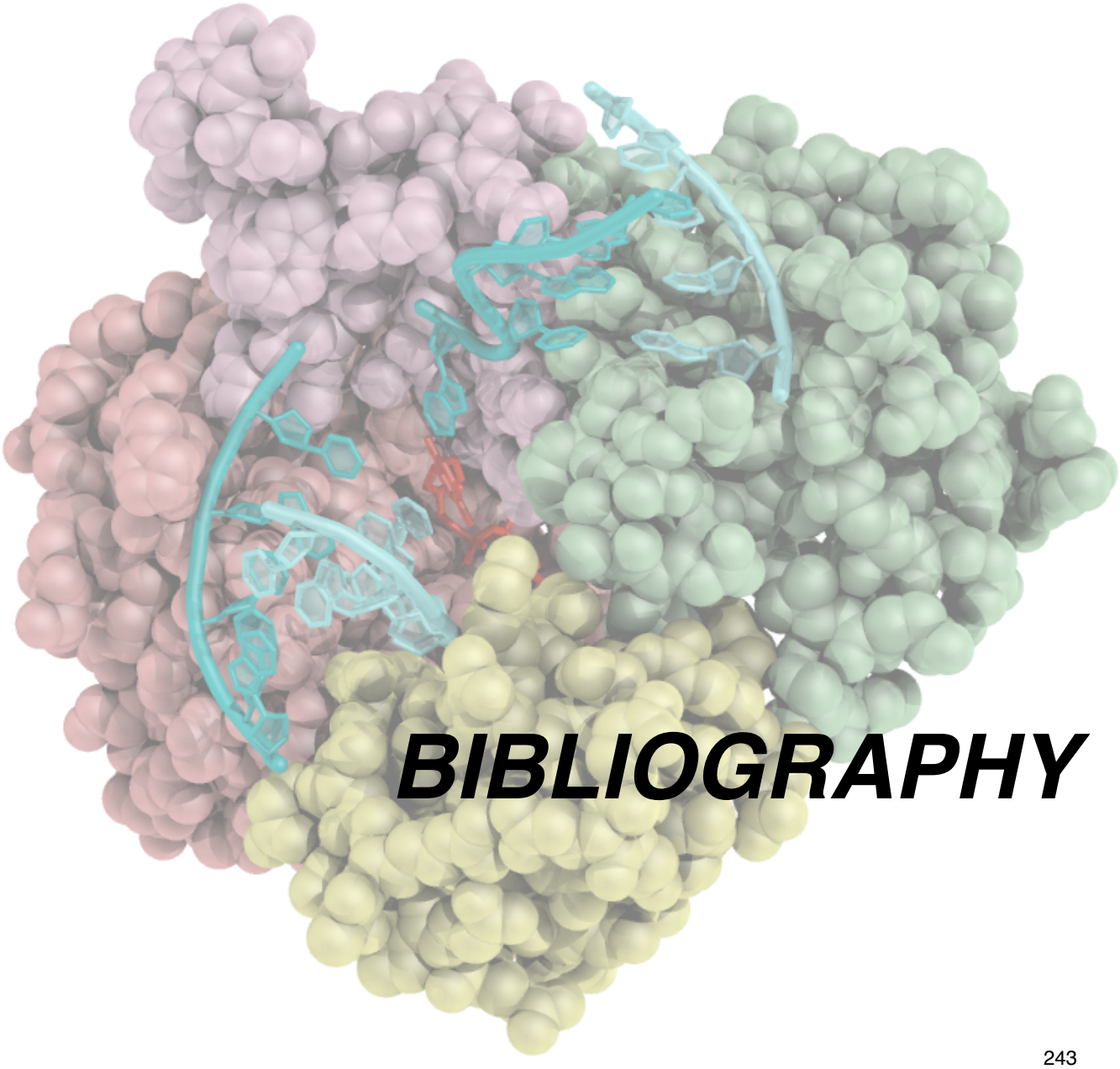
6. El Loop 1 es una pieza flexible de Pol*μ* que tiene un papel fundamental en la transferasa terminal y en la reunión de extremos de DNA, ya que actúa como un *pseudo-template* cuando la cadena molde es discontinua o no está disponible, mientras que se desorganiza cuando ésta está presente para evitar choques estéricos. Nuestros resultados de mutagénesis dirigida muestran que el "mini-loop" del dominio *thumb* (motivo NSH) y la hélice N, cargada positivamente, contribuyen en gran medida a la correcta colocación bien de la cadena molde o del Loop 1 en cada caso. Otras mutaciones en las "bisagras" del Loop 1, tales como los residuos $Phe^{385}$ o $Phe^{389}$, corroboran los requisitos de flexibilidad de este motivo.

7. Hemos identificado dos regiones con carga positiva en Pol*μ*, dedicadas específicamente a la unión de sustratos de NHEJ como entidades separadas: la zona de unión del *primer*, que contiene los residuos $Lys^{249}$, $Arg^{253}$ y $Arg^{416}$, y la zona de unión del *template*, formada principalmente por los residuos $Arg^{442}$, $Arg^{445}$ y $Arg^{449}$ de la hélice N. La mutación individual de cualquiera de estos residuos, que carece de importancia en la mayoría de los casos para la actividad de relleno de *gaps*, abolió completamente la actividad de reunión de extremos.

8. En cuanto a las posibles modificaciones post-traduccionales implicadas en la regulación de Pol*μ*, hemos identificado los residuos que podrían ser fosforilados *in vivo*. Nuestras evidencias sugieren que el principal evento de modificación afecta a la funcionalidad del Loop 1, a través de la fosforilación de la $Ser^{372}$. Por lo tanto, las actividades más características de Pol*μ* (es decir, la transferasa terminal, la unión de extremos no compatibles) serían “apagadas” en ciertas fases del ciclo celular (S y G2) en las que la promiscuidad de Pol*μ* podría ser perjudicial para la célula.

9. En *Mycobacterium tuberculosis* la vía de NHEJ incluye una enzima relacionada con la familia AEP (PolDom) como la principal polimerasa implicada. Hemos examinado un nuevo complejo que contiene un monómero de PolDom, un sustrato de DNA *template/downstream*, un NTP entrante y dos iones metálicos activadores, y nuestros resultados muestran que, incluso en ausencia de la cadena *primer*, este complejo es catalíticamente competente y beneficioso para la posterior reacción de reunión de extremos. Este complejo pre-ternario también mostró que un residuo altamente conservado, $Arg^{220}$, que se encuentra en el loop2 (específico de enzimas relacionadas con NHEJ), es el principal conmutador para encender o apagar el sitio activo a través de un contacto inactivante con uno de los aspárticos catalíticos.

10. La resolución de una segunda estructura de PolDom mostró por primera vez una sinapsis complementaria de dos extremos de DNA, mantenida por un dímero de PolDom. El análisis de los datos 3D y los resultados bioquímicos indican que el loop2 no sólo participa en la catálisis, sino también en la correcta posición del *primer*, a través de la $Lys^{217}$. Por otra parte, la selección correcta de la base molde durante el NHEJ se basa en la acción de dos fenilalaninas, que proveen a PolDom con la posibilidad ya sea de copiar correctamente el molde o de generar cambios en el marco de lectura (*frameshifts*), proporcionando así la versatilidad necesaria para llevar a cabo la reacción de reunión de extremos.



# BIBLIOGRAPHY

Abbotts, J, et al. (1988), 'Expression of human DNA polymerase beta in Escherichia coli and characterization of the recombinant enzyme.', *Biochemistry*, 27 (3), 901-09.

Ahn, J, BG Werneburg, and MD Tsai (1997), 'DNA polymerase beta: structure-fidelity relationship from Pre-steady-state kinetic analyses of all possible correct and incorrect base pairs for wild type and R283A mutant.', *Biochemistry*, 36 (5), 1100-07.

Ahnesorg, P, P Smith, and SP Jackson (2006), 'XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining.', *Cell*, 124 (2), 301-13.

Aleem, E, C Berthet, and P Kaldis (2004), 'Cdk2 as a master of S phase entry: fact or fake?', *Cell Cycle*, 3 (1), 35-37.

Alseth, I, et al. (2005), 'Biochemical characterization and DNA repair pathway interactions of Mag1-mediated base excision repair in Schizosaccharomyces pombe.', *Nucleic Acids Res*, 33 (3), 1123-31.

Andrade, P, et al. (2009), 'Limited terminal transferase in human DNA polymerase mu defines the required balance between accuracy and efficiency in NHEJ.', *Proc Natl Acad Sci U S A*, 106 (38), 16203-08.

Aoufouchi, S, et al. (2000), 'Two novel human and mouse DNA polymerases of the polX family.', *Nucleic Acids Res*, 28 (18), 3684-93.

Aravind, L and EV Koonin (2001), 'Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system.', *Genome Res*, 11 (8), 1365-74.

Arnold, K, et al. (2006), 'The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.', *Bioinformatics*, 22 (2), 195-201.

Astatke, M, ND Grindley, and CM Joyce (1998), 'How E. coli DNA polymerase I (Klenow fragment) distinguishes between deoxy- and dideoxynucleotides.', *J Mol Biol*, 278 (1), 147-65.

Beard, WA, et al. (1996), 'Enzyme-DNA interactions required for efficient nucleotide incorporation and discrimination in human DNA polymerase beta.', *J Biol Chem*, 271 (21), 12141-44.

Beard, WA and SH Wilson (1998), 'Structural insights into DNA polymerase beta fidelity: hold tight if you want it right.', *Chem Biol*, 5 (1), R7-13.

Beard, WA and SH Wilson (2000), 'Structural design of a eukaryotic DNA repair polymerase: DNA polymerase beta.', *Mutat Res*, 460 (3-4), 231-44.

Beard, WA, DD Shock, and SH Wilson (2004), 'Influence of DNA structure on DNA polymerase beta active site function: extension of mutagenic DNA intermediates.', *J Biol Chem*, 279 (30), 31921-29.

Beard, WA, R Prasad, and SH Wilson (2006), 'Activities and mechanism of DNA polymerase beta.', *Methods Enzymol*, 408 91-107.

Bebenek, K and TA Kunkel (1990), 'Frameshift errors initiated by nucleotide misincorporation.', *Proc Natl Acad Sci U S A*, 87 (13), 4946-50.

Bebenek, K, et al. (2001), '5'-Deoxyribose phosphate lyase activity of human DNA polymerase iota in vitro.', *Science*, 291 (5511), 2156-59.

Bebenek, K, et al. (2003), 'The frameshift infidelity of human DNA polymerase lambda. Implications for function.', *J Biol Chem*, 278 (36), 34685-90.

Bebenek, K and TA Kunkel (2004), 'Functions of DNA polymerases.', *Adv Protein Chem*, 69 137-65.

Bebenek, K, et al. (2005), 'Biochemical properties of Saccharomyces cerevisiae DNA polymerase IV.', *J Biol Chem*, 280 (20), 20051-58.

Bebenek, K, et al. (2010), 'Loop 1 modulates the fidelity of DNA polymerase lambda.', *Nucleic Acids Res*, 38 (16), 5419-31.

Beckman, KB and BN Ames (1997), 'Oxidative decay of DNA.', *J Biol Chem*, 272 (32), 19633-36.

Beese, LS and TA Steitz (1991), 'Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism.', *EMBO J*, 10 (1), 25-33.

Belousova, EA, et al. (2010), 'DNA polymerases beta and lambda bypass thymine glycol in gapped DNA structures.', *Biochemistry*, 49 (22), 4695-704.

Benkovic, SJ and CE Cameron (1995), 'Kinetic analysis of nucleotide incorporation and misincorporation by Klenow fragment of Escherichia coli DNA polymerase I.', *Methods Enzymol*, 262 257-69.

Bennett, RA, et al. (1997), 'Interaction of human apurinic endonuclease and DNA polymerase beta in the base excision repair pathway.', *Proc Natl Acad Sci U S A*, 94 (14), 7166-69.

Bentolila, LA, et al. (1995), 'The two isoforms of mouse terminal deoxynucleotidyl transferase differ in both the ability to add N regions and subcellular localization.', *EMBO J*, 14 (17), 4221-29.

Bentolila, LA, et al. (1999), 'Extensive junctional diversity in Ig light chain genes from early B cell progenitors of mu MT mice.', *J Immunol*, 162 (4), 2123-28.

Bernad, A, et al. (1989), 'A conserved 3'----5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases.', *Cell*, 59 (1), 219-28.

Bertocci, B, et al. (2003), 'Immunoglobulin kappa light chain gene rearrangement is impaired in mice deficient for DNA polymerase mu.', *Immunity*, 19 (2), 203-11.

Bertocci, B, et al. (2006), 'Nonoverlapping functions of DNA polymerases mu, lambda, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination in vivo.', *Immunity*, 25 (1), 31-41.

Biertumpfel, C, et al. (2010), 'Structure and mechanism of human DNA polymerase eta.', *Nature*, 465 (7301), 1044-48.

Blackwell, TK, et al. (1989), 'Isolation of scid pre-B cells that rearrange kappa light chain genes: formation of normal signal and abnormal coding joins.', *EMBO J*, 8 (3), 735-42.

Blanca, G, et al. (2003), 'Human DNA polymerase lambda diverged in evolution from DNA polymerase beta toward specific Mn(++) dependence: a kinetic and thermodynamic study.', *Biochemistry*, 42 (24), 7467-76.

Blanca, G, et al. (2004), 'Human DNA polymerases lambda and beta show different efficiencies of translesion DNA synthesis past abasic sites and alternative mechanisms for frameshift generation.', *Biochemistry*, 43 (36), 11605-15.

Blanco, L and M Salas (1984), 'Characterization and purification of a phage phi 29-encoded DNA polymerase required for the initiation of replication.', *Proc Natl Acad Sci U S A*, 81 (17), 5325-29.

Bloom, LB, et al. (1997), 'Fidelity of Escherichia coli DNA polymerase III holoenzyme. The effects of beta, gamma complex processivity proteins and epsilon proofreading exonuclease on nucleotide misincorporation efficiencies.', *J Biol Chem*, 272 (44), 27919-30.

Bollum, FJ (1959), 'Mammalian enzymes of desoxyribonucleic acid synthesis.', *Ann N Y Acad Sci*, 81 792-93.

Bollum, FJ (1960), 'Calf thymus polymerase.', *J Biol Chem*, 235 2399-403.

Bollum, FJ (1964), 'Chemically Defined Templates and Initiators for Deoxypolynucleotide Synthesis.', *Science*, 144 (3618), 560.

Bollum, FJ (1979), 'Terminal deoxynucleotidyl transferase as a hematopoietic cell marker.', *Blood*, 54 (6), 1203-15.

Bonnin, A, et al. (1999), 'A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type phi29 DNA polymerase.', *J Mol Biol*, 290 (1), 241-51.

Bork, P, et al. (1997), 'A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins.', *FASEB J*, 11 (1), 68-76.

Boule, JB, F Rougeon, and C Papanicolaou (2001), 'Terminal deoxynucleotidyl transferase indiscriminately incorporates ribonucleotides and deoxyribonucleotides.', *J Biol Chem*, 276 (33), 31388-93.

Boulton, SJ and SP Jackson (1998), 'Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing.', *EMBO J*, 17 (6), 1819-28.

Bowater, R and AJ Doherty (2006), 'Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining.', *PLoS Genet*, 2 (2), e8.

Braithwaite, DK and J Ito (1993), 'Compilation, alignment, and phylogenetic relationships of DNA polymerases.', *Nucleic Acids Res*, 21 (4), 787-802.

Braithwaite, EK, et al. (2005), 'DNA polymerase lambda mediates a back-up base excision repair activity in extracts of mouse embryonic fibroblasts.', *J Biol Chem*, 280 (18), 18469-75.

Brautigam, CA and TA Steitz (1998), 'Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes.', *Curr Opin Struct Biol*, 8 (1), 54-63.

Brissett, NC, et al. (2007), 'Structure of a NHEJ polymerase-mediated DNA synaptic complex.', *Science*, 318 (5849), 456-59.

Brissett, NC, et al. (2011), 'Structure of a preternary complex involving a prokaryotic NHEJ DNA polymerase.', *Mol Cell*, 41 (2), 221-31.

Brown, JA, et al. (2010), 'A novel mechanism of sugar selection utilized by a human X-family DNA polymerase.', *J Mol Biol*, 395 (2), 282-90.

Brown, JA and Z Suo (2011), 'Unlocking the sugar "steric gate" of DNA polymerases.', *Biochemistry*, 50 (7), 1135-42.

Bruton, RK, et al. (2007), 'C-terminal-binding protein interacting protein binds directly to adenovirus early region 1A through its N-terminal region and conserved region 3.', *Oncogene*, 26 (53), 7467-79.

Burgers, PM, et al. (2001), 'Eukaryotic DNA polymerases: proposal for a revised nomenclature.', *J Biol Chem*, 276 (47), 43487-90.

Cadet, J, et al. (1997), 'Oxidative damage to DNA: formation, measurement, and biological significance.', *Rev Physiol Biochem Pharmacol*, 131 1-87.

Caldecott, KW, et al. (1996), 'XRCC1 polypeptide interacts with DNA polymerase beta and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro.', *Nucleic Acids Res*, 24 (22), 4387-94.

Callebaut, I and JP Mornon (1997), 'From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair.', *FEBS Lett*, 400 (1), 25-30.

Castilla-Llorente, V, et al. (2006), 'Spo0A, the key transcriptional regulator for entrance into sporulation, is an inhibitor of DNA replication.', *EMBO J*, 25 (16), 3890-99.

Catalan, N, et al. (2003), 'The block in immunoglobulin class switch recombination caused by activation-induced cytidine deaminase deficiency occurs prior to the generation of DNA double strand breaks in switch mu region.', *J Immunol*, 171 (5), 2504-09.

Cavanaugh, NA, WA Beard, and SH Wilson (2010), 'DNA polymerase beta ribonucleotide discrimination: insertion, misinsertion, extension, and coding.', *J Biol Chem*, 285 (32), 24457-65.

Cavanaugh, NA, et al. (2011), 'Molecular insights into DNA polymerase deterrents for Ribonucleotide insertion.', *J Biol Chem*,

CCP4 (1994), 'The CCP4 suite: programs for protein crystallography.', *Acta Crystallogr D Biol Crystallogr*, 50 (Pt 5), 760-63.

Chagovetz, AM, JB Sweasy, and BD Preston (1997), 'Increased activity and fidelity of DNA polymerase beta on single-nucleotide gapped DNA.', *J Biol Chem*, 272 (44), 27501-04.

Chappell, C, et al. (2002), 'Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining.', *EMBO J*, 21 (11), 2827-32.
Chaudhuri, J and FW Alt (2004), 'Class-switch recombination: interplay of transcription, DNA deamination and DNA repair.', *Nat Rev Immunol*, 4 (7), 541-52.
Chen, D, et al. (2005), 'ARF-BP1/Mule is a critical mediator of the ARF tumor suppressor.', *Cell*, 121 (7), 1071-83.
Chen, L, et al. (2000), 'Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase.', *J Biol Chem*, 275 (34), 26196-205.
Chen, S, et al. (2001), 'Accurate in vitro end joining of a DNA double strand break with partially cohesive 3'-overhangs and 3'-phosphoglycolate termini: effect of Ku on repair fidelity.', *J Biol Chem*, 276 (26), 24323-30.
Coleman, MS, JJ Hutton, and FJ Bollum (1974), 'DNA polymerases in normal and leukemic human hematopoietic cells.', *Blood*, 44 (1), 19-32.
Corneo, B, et al. (2007), 'Rag mutations reveal robust alternative end joining.', *Nature*, 449 (7161), 483-86.
Covo, S, L Blanco, and Z Livneh (2004), 'Lesion bypass by human DNA polymerase mu reveals a template-dependent, sequence-independent nucleotidyl transferase activity.', *J Biol Chem*, 279 (2), 859-65.
Crespan, E, et al. (2007), 'Expanding the repertoire of DNA polymerase substrates: template-instructed incorporation of non-nucleoside triphosphate analogues by DNA polymerases beta and lambda.', *Nucleic Acids Res*, 35 (1), 45-57.
Crespan, E, U Hubscher, and G Maga (2007), 'Error-free bypass of 2-hydroxyadenine by human DNA polymerase lambda with Proliferating Cell Nuclear Antigen and Replication Protein A in different sequence contexts.', *Nucleic Acids Res*, 35 (15), 5173-81.
d'Adda di Fagagna, F, et al. (2001), 'Effects of DNA nonhomologous end-joining factors on telomere length and chromosomal stability in mammalian cells.', *Curr Biol*, 11 (15), 1192-96.
Daley, JM, et al. (2005), 'DNA joint dependence of pol X family polymerase action in nonhomologous end joining.', *J Biol Chem*, 280 (32), 29030-37.
Davis, BJ, JM Havener, and DA Ramsden (2008), 'End-bridging is required for pol mu to efficiently promote repair of noncomplementary ends by nonhomologous end joining.', *Nucleic Acids Res*, 36 (9), 3085-94.
de Laat, WL, NG Jaspers, and JH Hoeijmakers (1999), 'Molecular mechanism of nucleotide excision repair.', *Genes Dev*, 13 (7), 768-85.
DeFazio, LG, et al. (2002), 'Synapsis of DNA ends by DNA-dependent protein kinase.', *EMBO J*, 21 (12), 3192-200.
Deibel, MR Jr and MS Coleman (1980), 'Biochemical properties of purified human terminal deoxynucleotidyltransferase.', *J Biol Chem*, 255 (9), 4206-12.
Delarue, M, et al. (2002), 'Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase.', *EMBO J*, 21 (3), 427-39.
Della, M, et al. (2004), 'Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine.', *Science*, 306 (5696), 683-85.
Demogines, A, et al. (2010), 'Ancient and recent adaptive evolution of primate non-homologous end joining genes.', *PLoS Genet*, 6 (10), e1001169.
Derbyshire, V, ND Grindley, and CM Joyce (1991), 'The 3'-5' exonuclease of DNA polymerase I of Escherichia coli: contribution of each amino acid at the active site to the reaction.', *EMBO J*, 10 (1), 17-24.
DeRose, EF, et al. (2003), 'Solution structure of the lyase domain of human DNA polymerase lambda.', *Biochemistry*, 42 (32), 9564-74.
DeRose, EF, et al. (2007), 'Solution structure of polymerase mu's BRCT Domain reveals an element essential for its role in nonhomologous end joining.', *Biochemistry*, 46 (43), 12100-10.
Desiderio, SV, et al. (1984), 'Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells.', *Nature*, 311 (5988), 752-55.
Dianova, II, et al. (2004), 'XRCC1-DNA polymerase beta interaction is required for efficient base excision repair.', *Nucleic Acids Res*, 32 (8), 2550-55.
Dimitriadis, EK, et al. (1998), 'Thermodynamics of human DNA ligase I trimerization and association with DNA polymerase beta.', *J Biol Chem*, 273 (32), 20540-50.
Dogliotti, E, et al. (2001), 'The mechanism of switching among multiple BER pathways.', *Prog Nucleic Acid Res Mol Biol*, 68 3-27.
Doherty, AJ, LC Serpell, and CP Ponting (1996), 'The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA.', *Nucleic Acids Res*, 24 (13), 2488-97.
Dominguez, O, et al. (2000), 'DNA polymerase mu (Pol mu), homologous to TdT, could act as a DNA mutator in eukaryotic cells.', *EMBO J*, 19 (7), 1731-42.
Doublie, S, MR Sawaya, and T Ellenberger (1999), 'An open and closed case for all polymerases.', *Structure*, 7 (2), R31-5.
Dudley, DD, et al. (2005), 'Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences.', *Adv Immunol*, 86 43-112.
Dunnick, W, et al. (1993), 'DNA sequences at immunoglobulin switch region recombination sites.', *Nucleic Acids Res*, 21 (3), 365-72.
Duvauchelle, JB, et al. (2002), 'Human DNA polymerase mu (Pol mu) exhibits an unusual replication slippage ability at AAF lesion.', *Nucleic Acids Res*, 30 (9), 2061-67.

Dvir, A, et al. (1992), 'Ku autoantigen is the regulatory component of a template-associated protein kinase that phosphorylates RNA polymerase II.', *Proc Natl Acad Sci U S A*, 89 (24), 11920-24.
Dynan, WS and S Yoo (1998), 'Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids.', *Nucleic Acids Res*, 26 (7), 1551-59.
Edwards, SL, et al. (2008), 'Resistance to therapy caused by intragenic deletion in BRCA2.', *Nature*, 451 (7182), 1111-15.
Efrati, E, et al. (1997), 'Abasic translesion synthesis by DNA polymerase beta violates the "A-rule". Novel types of nucleotide incorporation by human DNA polymerase beta at an abasic lesion in different sequence contexts.', *J Biol Chem*, 272 (4), 2559-69.
El-Andaloussi, N, et al. (2006), 'Arginine methylation regulates DNA polymerase beta.', *Mol Cell*, 22 (1), 51-62.
El-Andaloussi, N, et al. (2007), 'Methylation of DNA polymerase beta by protein arginine methyltransferase 1 regulates its binding to proliferating cell nuclear antigen.', *FASEB J*, 21 (1), 26-34.
Emsley, P and K Cowtan (2004), 'Coot: model-building tools for molecular graphics.', *Acta Crystallogr D Biol Crystallogr*, 60 (Pt 12 Pt 1), 2126-32.
Espejel, S, et al. (2004), 'Shorter telomeres, accelerated ageing and increased lymphoma in DNA-PKcs-deficient mice.', *EMBO Rep*, 5 (5), 503-09.
Evans, JD and P Hearing (2005), 'Relocalization of the Mre11-Rad50-Nbs1 complex by the adenovirus E4 ORF3 protein is required for viral replication.', *J Virol*, 79 (10), 6207-15.
Fan, W and X Wu (2004), 'DNA polymerase lambda can elongate on DNA substrates mimicking non-homologous end joining and interact with XRCC4-ligase IV complex.', *Biochem Biophys Res Commun*, 323 (4), 1328-33.
Feldmann, E, et al. (2000), 'DNA double-strand break repair in cell-free extracts from Ku80-deficient cells: implications for Ku serving as an alignment factor in non-homologous DNA end joining.', *Nucleic Acids Res*, 28 (13), 2585-96.
Ferguson, DO, et al. (2000), 'The nonhomologous end-joining pathway of DNA repair is required for genomic stability and the suppression of translocations.', *Proc Natl Acad Sci U S A*, 97 (12), 6630-33.
Ferraro, P, et al. (1994), 'The CCP4 suite: programs for protein crystallography.', *Acta Crystallogr D Biol Crystallogr*, 50 (Pt 5), 760-63.
Ferraro, P, et al. (2010), 'Quantitation of cellular deoxynucleoside triphosphates.', *Nucleic Acids Res*, 38 (6), e85.
Fiala, KA, et al. (2006), 'Up-regulation of the fidelity of human DNA polymerase lambda by its non-enzymatic proline-rich domain.', *J Biol Chem*, 281 (28), 19038-44.
Fidalgo da Silva, E, SS Mandal, and LJ Reha-Krantz (2002), 'Using 2-aminopurine fluorescence to measure incorporation of incorrect nucleotides by wild type and mutant bacteriophage T4 DNA polymerases.', *J Biol Chem*, 277 (43), 40640-49.
Foley, MC, K Arora, and T Schlick (2006), 'Sequential side-chain residue motions transform the binary into the ternary state of DNA polymerase lambda.', *Biophys J*, 91 (9), 3182-95.
Foley, MC and T Schlick (2009), 'Relationship between conformational changes in pol lambda's active site upon binding incorrect nucleotides and mismatch incorporation rates.', *J Phys Chem B*, 113 (39), 13035-47.
Fortini, P, et al. (1998), 'Different DNA polymerases are involved in the short- and long-patch base excision repair in mammalian cells.', *Biochemistry*, 37 (11), 3575-80.
Frank, EG and R Woodgate (2007), 'Increased catalytic activity and altered fidelity of human DNA polymerase iota in the presence of manganese.', *J Biol Chem*, 282 (34), 24689-96.
Frank, KM, et al. (1998), 'Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV.', *Nature*, 396 (6707), 173-77.
Franklin, WA and T Lindahl (1988), 'DNA deoxyribophosphodiesterase.', *EMBO J*, 7 (11), 3617-22.
Freemont, PS, et al. (1988), 'Cocrystal structure of an editing complex of Klenow fragment with DNA.', *Proc Natl Acad Sci U S A*, 85 (23), 8924-28.
Frick, DN and CC Richardson (2001), 'DNA primases.', *Annu Rev Biochem*, 70 39-80.
Frouin, I, et al. (2005), 'Phosphorylation of human DNA polymerase lambda by the cyclin-dependent kinase Cdk2/cyclin A complex is modulated by its association with proliferating cell nuclear antigen.', *Nucleic Acids Res*, 33 (16), 5354-61.
Gao, Y, et al. (1998), 'A targeted DNA-PKcs-null mutation reveals DNA-PK-independent functions for KU in V(D)J recombination.', *Immunity*, 9 (3), 367-76.
Garcia-Diaz, M, et al. (2000), 'DNA polymerase lambda (Pol lambda), a novel eukaryotic DNA polymerase with a potential role in meiosis.', *J Mol Biol*, 301 (4), 851-67.
Garcia-Diaz, M, et al. (2001), 'Identification of an intrinsic 5'-deoxyribose-5-phosphate lyase activity in human DNA polymerase lambda: a possible role in base excision repair.', *J Biol Chem*, 276 (37), 34659-63.
Garcia-Diaz, M, et al. (2002), 'DNA polymerase lambda, a novel DNA repair enzyme in human cells.', *J Biol Chem*, 277 (15), 13184-91.
Garcia-Diaz, M, et al. (2004), 'A structural solution for the DNA polymerase lambda-dependent repair of DNA gaps with minimal homology.', *Mol Cell*, 13 (4), 561-72.
Garcia-Diaz, M, et al. (2005), 'Structure-function studies of DNA polymerase lambda.', *DNA Repair (Amst)*, 4 (12), 1358-67.
Garcia-Diaz, M, et al. (2005), 'A closed conformation for the Pol lambda catalytic cycle.', *Nat Struct Mol Biol*, 12 (1), 97-98.

Garcia-Diaz, M and TA Kunkel (2006), 'Mechanism of a genetic glissando: structural biology of indel mutations.', *Trends Biochem Sci*, 31 (4), 206-14.
Garcia-Diaz, M, et al. (2007), 'Role of the catalytic metal during polymerization by DNA polymerase lambda.', *DNA Repair (Amst)*, 6 (9), 1333-40.
Garcia-Diaz, M, et al. (2009), 'Template strand scrunching during DNA gap repair synthesis by human polymerase lambda.', *Nat Struct Mol Biol*, 16 (9), 967-72.
Garcia-Escudero, R, et al. (2003), 'DNA polymerase X of African swine fever virus: insertion fidelity on gapped DNA substrates and AP lyase activity support a role in base excision repair of viral DNA.', *J Mol Biol*, 326 (5), 1403-12.
Geng, Y, et al. (2003), 'Cyclin E ablation in the mouse.', *Cell*, 114 (4), 431-43.
Gong, C, et al. (2004), 'Biochemical and genetic analysis of the four DNA ligases of mycobacteria.', *J Biol Chem*, 279 (20), 20594-606.
Gong, C, et al. (2005), 'Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C.', *Nat Struct Mol Biol*, 12 (4), 304-12.
Gonzalez-Barrera, S, et al. (2005), 'Characterization of SpPol4, a unique X-family DNA polymerase in Schizosaccharomyces pombe.', *Nucleic Acids Res*, 33 (15), 4762-74.
Goodman, MF and B Tippin (2000), 'The expanding polymerase universe.', *Nat Rev Mol Cell Biol*, 1 (2), 101-09.
Goodman, MF (2002), 'Error-prone repair DNA polymerases in prokaryotes and eukaryotes.', *Annu Rev Biochem*, 71 17-50.
Gotte, M (2006), 'Effects of nucleotides and nucleotide analogue inhibitors of HIV-1 reverse transcriptase in a ratchet model of polymerase translocation.', *Curr Pharm Des*, 12 (15), 1867-77.
Gottlieb, TM and SP Jackson (1993), 'The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen.', *Cell*, 72 (1), 131-42.
Gozalbo-Lopez, B, et al. (2009), 'A role for DNA polymerase mu in the emerging DJH rearrangements of the postgastrulation mouse embryo.', *Mol Cell Biol*, 29 (5), 1266-75.
Grawunder, U, et al. (1997), 'Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells.', *Nature*, 388 (6641), 492-95.
Greenman, C, et al. (2007), 'Patterns of somatic mutation in human cancer genomes.', *Nature*, 446 (7132), 153-58.
Gryk, MR, et al. (2002), 'Mapping of the interaction interface of DNA polymerase beta with XRCC1.', *Structure*, 10 (12), 1709-20.
Gu, Y, et al. (1997), 'Ku70-deficient embryonic stem cells have increased ionizing radiosensitivity, defective DNA end-binding activity, and inability to support V(D)J recombination.', *Proc Natl Acad Sci U S A*, 94 (15), 8076-81.
Guirouilh-Barbat, J, et al. (2004), 'Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells.', *Mol Cell*, 14 (5), 611-23.
Guirouilh-Barbat, J, et al. (2007), 'Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends.', *Proc Natl Acad Sci U S A*, 104 (52), 20902-07.
Hande, P, et al. (1999), 'Elongated telomeres in scid mice.', *Genomics*, 56 (2), 221-23.
Hanes, JW, et al. (2007), 'Enzymatic therapeutic index of acyclovir. Viral versus human polymerase gamma specificity.', *J Biol Chem*, 282 (34), 25159-67.
Hanes, JW and KA Johnson (2008), 'Real-time measurement of pyrophosphate release kinetics.', *Anal Biochem*, 372 (1), 125-27.
Hartlerode, AJ and R Scully (2009), 'Mechanisms of double-strand break repair in somatic mammalian cells.', *Biochem J*, 423 (2), 157-68.
Hasan, S, et al. (2002), 'Acetylation regulates the DNA end-trimming activity of DNA polymerase beta.', *Mol Cell*, 10 (5), 1213-22.
Havener, JM, et al. (2003), 'Translesion synthesis past platinum DNA adducts by human DNA polymerase mu.', *Biochemistry*, 42 (6), 1777-88.
Hefferin, ML and AE Tomkinson (2005), 'Mechanism of DNA double-strand break repair by non-homologous end joining.', *DNA Repair (Amst)*, 4 (6), 639-48.
Heidenreich, E, et al. (2003), 'Non-homologous end joining as an important mutagenic process in cell cycle-arrested cells.', *EMBO J*, 22 (9), 2274-83.
Helbock, HJ, et al. (1998), 'DNA oxidation matters: the HPLC-electrochemical detection assay of 8-oxo-deoxyguanosine and 8-oxo-guanine.', *Proc Natl Acad Sci U S A*, 95 (1), 288-93.
Hentges, P, et al. (2006), 'Evolutionary and functional conservation of the DNA non-homologous end-joining protein, XLF/Cernunnos.', *J Biol Chem*, 281 (49), 37517-26.
Hirose, F, et al. (1989), 'Difference in the expression level of DNA polymerase beta among mouse tissues: high expression in the pachytene spermatocyte.', *Exp Cell Res*, 181 (1), 169-80.
Hoege, C, et al. (2002), 'RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO.', *Nature*, 419 (6903), 135-41.
Hoeijmakers, JH (2001), 'DNA repair mechanisms.', *Maturitas*, 38 (1), 17-22; discussion 22-3.
Holm, L and C Sander (1995), 'DNA polymerase beta belongs to an ancient nucleotidyltransferase superfamily.', *Trends Biochem Sci*, 20(9) (9), 345-47.
Hubscher, U, G Maga, and S Spadari (2002), 'Eukaryotic DNA polymerases.', *Annu Rev Biochem*, 71 133-63.

Huen, MS and J Chen (2008), 'The DNA damage response pathways: at the crossroad of protein modifications.', *Cell Res*, 18 (1), 8-16.
Idriss, HT, O Al-Assar, and SH Wilson (2002), 'DNA polymerase beta.', *Int J Biochem Cell Biol*, 34 (4), 321-24.
Ito, J and DK Braithwaite (1991), 'Compilation and alignment of DNA polymerase sequences.', *Nucleic Acids Res*, 19 (15), 4045-57.
Iwasato, T, et al. (1990), 'Circular DNA is excised by immunoglobulin class switch recombination.', *Cell*, 62 (1), 143-49.
Iyer, LM, et al. (2005), 'Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members.', *Nucleic Acids Res*, 33 (12), 3875-96.
Jayaram, S, et al. (2008), 'E1B 55k-independent dissociation of the DNA ligase IV/XRCC4 complex by E4 34k during adenovirus infection.', *Virology*, 382 (2), 163-70.
Johnson, KA (2010), 'The kinetic and chemical mechanism of high-fidelity DNA polymerases.', *Biochim Biophys Acta*, 1804 (5), 1041-48.
Johnson, RE, et al. (2003), 'Deoxynucleotide triphosphate binding mode conserved in Y family DNA polymerases.', *Mol Cell Biol*, 23 (8), 3008-12.
Joyce, CM (1997), 'Choosing the right sugar: how polymerases select a nucleotide substrate.', *Proc Natl Acad Sci U S A*, 94 (5), 1619-22.
Joyce, CM (2004), 'T4 replication: what does "processivity" really mean?', *Proc Natl Acad Sci U S A*, 101 (22), 8255-56.
Juarez, R, et al. (2006), 'A specific loop in human DNA polymerase mu allows switching between creative and DNA-instructed synthesis.', *Nucleic Acids Res*, 34 (16), 4572-82.
Kaguni, LS (2004), 'DNA polymerase gamma, the mitochondrial replicase.', *Annu Rev Biochem*, 73 293-320.
Karanjawala, ZE, et al. (2002), 'Oxygen metabolism causes chromosome breaks and is associated with the neuronal apoptosis observed in DNA double-strand break repair mutants.', *Curr Biol*, 12 (5), 397-402.
Karimi-Busheri, F, et al. (1999), 'Molecular characterization of a human DNA kinase.', *J Biol Chem*, 274 (34), 24187-94.
Kato, KI, et al. (1967), 'Deoxynucleotide-polymerizing enzymes of calf thymus gland. II. Properties of the terminal deoxynucleotidyltransferase.', *J Biol Chem*, 242 (11), 2780-89.
Kelley, JL, et al. (2009), 'Targeted resequencing of two genes, RAGE and Polλ, confirms findings from a genome-wide scan for adaptive evolution and provides evidence for positive selection in additional populations.', *Hum Mol Genet*, 18 (4), 779-84.
Kelley, LA and MJ Sternberg (2009), 'Protein structure prediction on the Web: a case study using the Phyre server.', *Nat Protoc*, 4 (3), 363-71.
Kelman, Z (1997), 'PCNA: structure, functions and interactions.', *Oncogene*, 14 (6), 629-40.
Kiefer, F, et al. (2009), 'The SWISS-MODEL Repository and associated resources.', *Nucleic Acids Res*, 37 (Database issue), D387-92.
Kilzer, JM, et al. (2003), 'Roles of host cell factors in circularization of retroviral dna.', *Virology*, 314 (1), 460-67.
Kimura, S and K Sakaguchi (2006), 'DNA repair in plants.', *Chem Rev*, 106 (2), 753-66.
Komori, T, et al. (1993), 'Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes.', *Science*, 261 (5125), 1171-75.
Kotake, M, et al. (2002), 'Hormonal regulation of DNA polymerase beta activity and expression in rat adrenal glands and testes.', *Mol Cell Endocrinol*, 192 (1-2), 127-32.
Kouchakdjian, M, et al. (1991), 'NMR structural studies of the ionizing radiation adduct 7-hydro-8-oxodeoxyguanosine (8-oxo-7H-dG) opposite deoxyadenosine in a DNA duplex. 8-Oxo-7H-dG(syn).dA(anti) alignment at lesion site.', *Biochemistry*, 30 (5), 1403-12.
Krahn, JM, et al. (2003), 'Structure of DNA polymerase beta with the mutagenic DNA lesion 8-oxodeoxyguanine reveals structural insights into its coding potential.', *Structure*, 11 (1), 121-27.
Kramer, KM, et al. (1994), 'Two different types of double-strand breaks in Saccharomyces cerevisiae are repaired by similar RAD52-independent, nonhomologous recombination events.', *Mol Cell Biol*, 14 (2), 1293-301.
Krayevsky, AA, et al. (2000), 'Terminal deoxynucleotidyl transferase. catalysis of DNA (oligodeoxynucleotide) phosphorylation.', *Pharmacol Ther*, 85 (3), 165-73.
Krejci, L, et al. (2003), 'Mending the break: two DNA double-strand break repair machines in eukaryotes.', *Prog Nucleic Acid Res Mol Biol*, 74 159-201.
Kubota, Y, et al. (1996), 'Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein.', *EMBO J*, 15 (23), 6662-70.
Kulaeva, OI, et al. (1996), 'Identification of a DinB/UmuC homolog in the archeon Sulfolobus solfataricus.', *Mutat Res*, 357 (1-2), 245-53.
Kumar, A, et al. (1990), 'Studies of the domain structure of mammalian DNA polymerase beta. Identification of a discrete template binding domain.', *J Biol Chem*, 265 (4), 2124-31.
Kumar, S, M Bakhtina, and MD Tsai (2008), 'Altered order of substrate binding by DNA polymerase X from African Swine Fever virus.', *Biochemistry*, 47 (30), 7875-87.
Kunkel, TA (1985), 'The mutational specificity of DNA polymerase-beta during in vitro DNA synthesis. Production of frameshift, base substitution, and deletion mutations.', *J Biol Chem*, 260 (9), 5787-96.
Kunkel, TA, et al. (1986), 'Rearrangements of DNA mediated by terminal transferase.', *Proc Natl Acad Sci U S A*, 83 (6), 1867-71.
Kunkel, TA (1988), 'Exonucleolytic proofreading.', *Cell*, 53 (6), 837-40.

Kunkel, TA and K Bebenek (2000), 'DNA replication fidelity.', *Annu Rev Biochem*, 69 497-529.
Kunkel, TA (2004), 'DNA replication fidelity.', *J Biol Chem*, 279 (17), 16895-98.
Lam, WC, et al. (1998), 'Effects of mutations on the partitioning of DNA substrates between the polymerase and 3'-5' exonuclease sites of DNA polymerase I (Klenow fragment).', *Biochemistry*, 37 (6), 1513-22.
Lao-Sirieix, SH, L Pellegrini, and SD Bell (2005), 'The promiscuous primase.', *Trends Genet*, 21 (10), 568-72.
Lee, JW, et al. (2003), 'Requirement for XRCC4 and DNA ligase IV in alignment-based gap filling for nonhomologous DNA end joining in vitro.', *Cancer Res*, 63 (1), 22-24.
Lee, JW, et al. (2004), 'Implication of DNA polymerase lambda in alignment-based gap filling for nonhomologous DNA end joining in human nuclear extracts.', *J Biol Chem*, 279 (1), 805-11.
Lehman, IR, et al. (1958), 'Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli.', *J Biol Chem*, 233 (1), 163-70.
Li, L, et al. (2001), 'Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection.', *EMBO J*, 20 (12), 3272-81.
Li, Y, et al. (1998), 'Crystal structures of the Klenow fragment of Thermus aquaticus DNA polymerase I complexed with deoxyribonucleoside triphosphates.', *Protein Sci*, 7 (5), 1116-23.
Li, Y and T Schlick (2010), 'Modeling DNA polymerase mu motions: subtle transitions before chemistry.', *Biophys J*, 99 (10), 3463-72.
Lieber, MR, et al. (1988), 'Lymphoid V(D)J recombination: nucleotide insertion at signal joints as well as coding joints.', *Proc Natl Acad Sci U S A*, 85 (22), 8588-92.
Lieber, MR, et al. (2003), 'Mechanism and regulation of human non-homologous DNA end-joining.', *Nat Rev Mol Cell Biol*, 4 (9), 712-20.
Lieber, MR, et al. (2004), 'The mechanism of vertebrate nonhomologous DNA end joining and its role in V(D)J recombination.', *DNA Repair (Amst)*, 3 (8-9), 817-26.
Lieber, MR (2008), 'The mechanism of human nonhomologous DNA end joining.', *J Biol Chem*, 283 (1), 1-5.
Lin, CW and A Engelman (2003), 'The barrier-to-autointegration factor is a component of functional human immunodeficiency virus type 1 preintegration complexes.', *J Virol*, 77 (8), 5030-36.
Lindahl, T, et al. (1993), 'DNA joining in mammalian cells.', *Cold Spring Harb Symp Quant Biol*, 58 619-24.
Lindahl, T, et al. (1997), 'Repair and processing events at DNA ends.', *Ciba Found Symp*, 211 198-205; discussion 205-8.
Ling, H, et al. (2001), 'Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication.', *Cell*, 107 (1), 91-102.
Ling, H, et al. (2003), 'Replication of a cis-syn thymine dimer at atomic resolution.', *Nature*, 424 (6952), 1083-87.
Ling, H, et al. (2004), 'Snapshots of replication through an abasic lesion; structural basis for base substitutions and frameshifts.', *Mol Cell*, 13 (5), 751-62.
Lipscomb, LA, et al. (1995), 'X-ray structure of a DNA decamer containing 7,8-dihydro-8-oxoguanine.', *Proc Natl Acad Sci U S A*, 92 (3), 719-23.
Lone, S, et al. (2007), 'Human DNA polymerase kappa encircles DNA: implications for mismatch extension and lesion bypass.', *Mol Cell*, 25 (4), 601-14.
Longley, MJ, et al. (1998), 'Identification of 5'-deoxyribose phosphate lyase activity in human DNA polymerase gamma and its role in mitochondrial base excision repair in vitro.', *Proc Natl Acad Sci U S A*, 95 (21), 12244-48.
Lucas, D, et al. (2005), 'Polymerase mu is up-regulated during the T cell-dependent immune response and its deficiency alters developmental dynamics of spleen centroblasts.', *Eur J Immunol*, 35 (5), 1601-11.
Lukacsovich, T, D Yang, and AS Waldman (1994), 'Repair of a specific double-strand break generated within a mammalian chromosome by yeast endonuclease I-SceI.', *Nucleic Acids Res*, 22 (25), 5649-57.
Luo, Y, et al. (2007), 'Impaired DNA repair via the base-excision repair pathway after focal ischemic brain injury: a protein phosphorylation-dependent mechanism reversed by hypothermic neuroprotection.', *Front Biosci*, 12 1852-62.
Ma, JL, et al. (2003), 'Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences.', *Mol Cell Biol*, 23 (23), 8820-28.
Ma, Y, et al. (2002), 'Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination.', *Cell*, 108 (6), 781-94.
Ma, Y, et al. (2004), 'A biochemically defined system for mammalian nonhomologous DNA end joining.', *Mol Cell*, 16 (5), 701-13.
Maciejewski, MW, et al. (2001), 'Solution structure of a viral DNA repair polymerase.', *Nat Struct Biol*, 8 (11), 936-41.
Maezawa, S, et al. (2008), 'Bood POZ containing gene type 2 is a human counterpart of yeast Btb3p and promotes the degradation of terminal deoxynucleotidyltransferase.', *Genes Cells*, 13 (5), 439-57.
Maga, G, et al. (2007), '8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins.', *Nature*, 447 (7144), 606-08.
Mahajan, KN, et al. (1999), 'Association of terminal deoxynucleotidyl transferase with Ku.', *Proc Natl Acad Sci U S A*, 96 (24), 13926-31.
Mahajan, KN, et al. (2002), 'Association of DNA polymerase mu (pol mu) with Ku and ligase IV: role for pol mu in end-joining double-strand break repair.', *Mol Cell Biol*, 22 (14), 5194-202.
Mahaney, BL, K Meek, and SP Lees-Miller (2009), 'Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining.', *Biochem J*, 417 (3), 639-50.

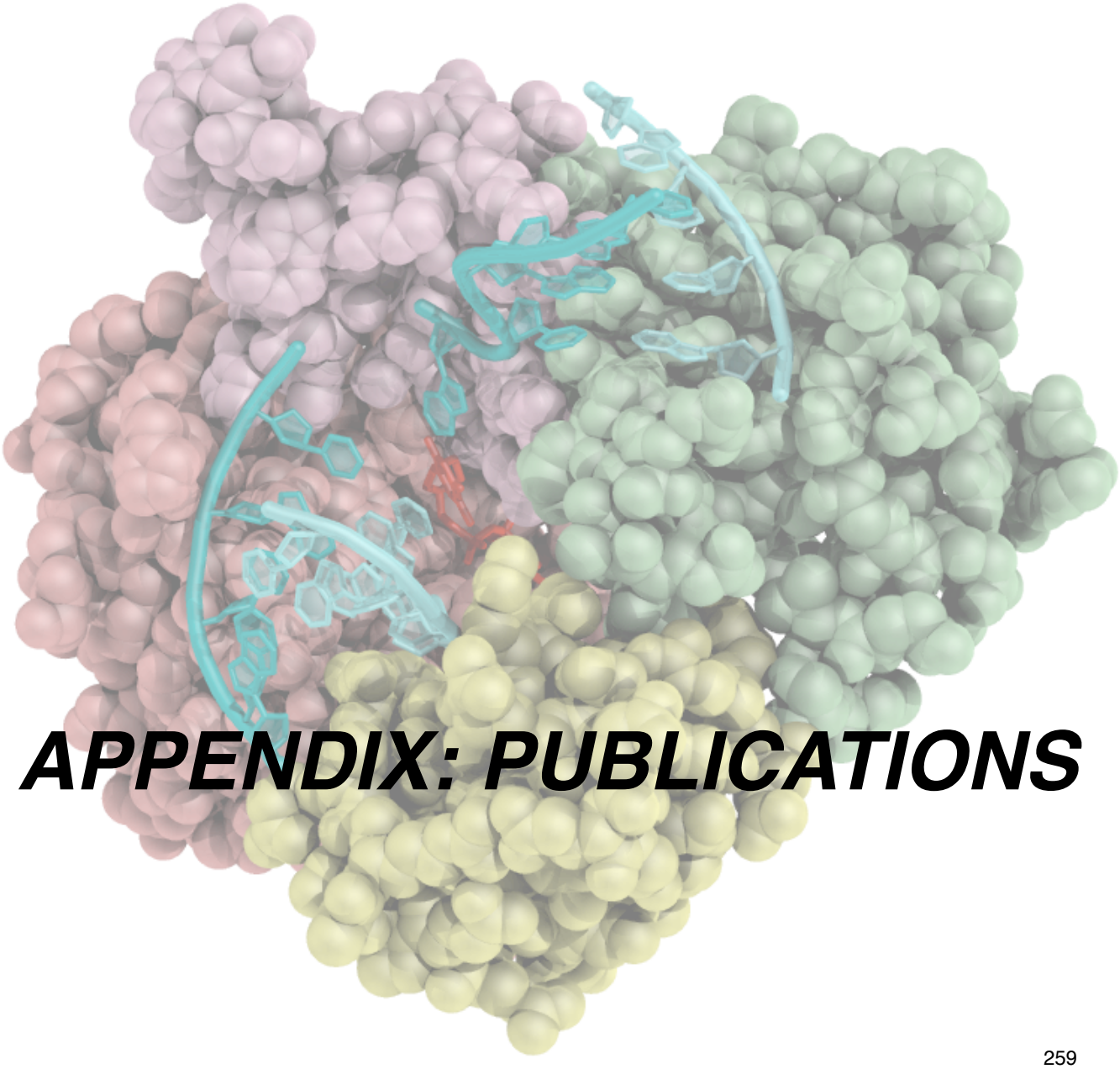
Manis, JP, et al. (1998), 'Ku70 is required for late B cell development and immunoglobulin heavy chain class switching.', *J Exp Med*, 187 (12), 2081-89.
Manis, JP, et al. (2002), 'IgH class switch recombination to IgG1 in DNA-PKcs-deficient B cells.', *Immunity*, 16 (4), 607-17.
Manke, IA, et al. (2003), 'BRCT repeats as phosphopeptide-binding modules involved in protein targeting.', *Science*, 302 (5645), 636-39.
Marini, F, et al. (2003), 'POLN, a nuclear PolA family DNA polymerase homologous to the DNA cross-link sensitivity protein Mus308.', *J Biol Chem*, 278 (34), 32014-19.
Marintchev, A, et al. (2000), 'Domain specific interaction in the XRCC1-DNA polymerase beta complex.', *Nucleic Acids Res*, 28 (10), 2049-59.
McAuley-Hecht, KE, et al. (1994), 'Crystal structure of a DNA duplex containing 8-hydroxydeoxyguanine-adenine base pairs.', *Biochemistry*, 33 (34), 10266-70.
McCoy, AJ, et al. (2005), 'Likelihood-enhanced fast translation functions.', *Acta Crystallogr D Biol Crystallogr*, 61 (Pt 4), 458-64.
Michel, B, SD Ehrlich, and M Uzest (1997), 'DNA double-strand breaks caused by replication arrest.', *EMBO J*, 16 (2), 430-38.
Miller, H, et al. (2000), '8-oxodGTP incorporation by DNA polymerase beta is modified by active-site residue Asn279.', *Biochemistry*, 39 (5), 1029-33.
Mills, KD, DO Ferguson, and FW Alt (2003), 'The role of DNA breaks in genomic instability and tumorigenesis.', *Immunol Rev*, 194 77-95.
Mombaerts, P, et al. (1992), 'RAG-1-deficient mice have no mature B and T lymphocytes.', *Cell*, 68 (5), 869-77.
Moon, AF, et al. (2007), 'Structural insight into the substrate specificity of DNA Polymerase mu.', *Nat Struct Mol Biol*, 14 (1), 45-53.
Moore, JK and JE Haber (1996), 'Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae.', *Mol Cell Biol*, 16 (5), 2164-73.
Moreno, S, A Klar, and P Nurse (1991), 'Molecular genetic analysis of fission yeast Schizosaccharomyces pombe.', *Methods Enzymol*, 194 795-823.
Morgan, DO (1997), 'Cyclin-dependent kinases: engines, clocks, and microprocessors.', *Annu Rev Cell Dev Biol*, 13 261-91.
Morgan, WF, et al. (1996), 'Genomic instability induced by ionizing radiation.', *Radiat Res*, 146 (3), 247-58.
Moriya, M (1993), 'Single-stranded shuttle phagemid for mutagenesis studies in mammalian cells: 8-oxoguanine in DNA induces targeted G.C-->T.A transversions in simian kidney cells.', *Proc Natl Acad Sci U S A*, 90 (3), 1122-26.
Mueller, GA, et al. (2008), 'A comparison of BRCT domains involved in nonhomologous end-joining: introducing the solution structure of the BRCT domain of polymerase lambda.', *DNA Repair (Amst)*, 7 (8), 1340-51.
Mullen, GP and SH Wilson (1997), 'DNA polymerase beta in abasic site repair: a structurally conserved helix-hairpin-helix motif in lesion detection by base excision repair enzymes.', *Biochemistry*, 36 (16), 4713-17.
Murshudov, GN, et al. (2011), 'REFMAC5 for the refinement of macromolecular crystal structures.', *Acta Crystallogr D Biol Crystallogr*, 67 (Pt 4), 355-67.
Myung, K, et al. (2004), 'Regulation of telomere length and suppression of genomic instability in human somatic cells by Ku86.', *Mol Cell Biol*, 24 (11), 5050-59.
Nagasawa, K, et al. (2000), 'Identification and characterization of human DNA polymerase beta 2, a DNA polymerase beta -related enzyme.', *J Biol Chem*, 275 (40), 31233-38.
Nakamura, J and JA Swenberg (1999), 'Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues.', *Cancer Res*, 59 (11), 2522-26.
Nelson, JR, CW Lawrence, and DC Hinkle (1996), 'Thymine-thymine dimer bypass by yeast DNA polymerase zeta.', *Science*, 272 (5268), 1646-49.
Nick McElhinny, SA and DA Ramsden (2003), 'Polymerase mu is a DNA-directed DNA/RNA polymerase.', *Mol Cell Biol*, 23 (7), 2309-15.
Nick McElhinny, SA, et al. (2005), 'A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining.', *Mol Cell*, 19 (3), 357-66.
Nick McElhinny, SA, et al. (2010), 'Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases.', *Proc Natl Acad Sci U S A*, 107 (11), 4949-54.
Nick McElhinny, SA, et al. (2010), 'Genome instability due to ribonucleotide incorporation into DNA.', *Nat Chem Biol*, 6 (10), 774-81.
Nicolas, N, et al. (1998), 'A human severe combined immunodeficiency (SCID) condition with increased sensitivity to ionizing radiations and impaired V(D)J rearrangements defines a new DNA recombination/repair deficiency.', *J Exp Med*, 188 (4), 627-34.
Oda, Y, et al. (1991), 'NMR studies for identification of dI:dG mismatch base-pairing structure in DNA.', *Nucleic Acids Res*, 19 (19), 5263-67.
Oettinger, MA, et al. (1990), 'RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination.', *Science*, 248 (4962), 1517-23.
Ohmori, H, et al. (2001), 'The Y-family of DNA polymerases.', *Mol Cell*, 8(1) (1), 7-8.

Oliveros, M, et al. (1997), 'Characterization of an African swine fever virus 20-kDa DNA polymerase involved in DNA repair.', *J Biol Chem*, 272 (49), 30899-910.
Ollis, DL, et al. (1985), 'Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP.', *Nature*, 313 (6005), 762-66.
Ortega, S, et al. (2003), 'Cyclin-dependent kinase 2 is essential for meiosis but not for mitotic cell division in mice.', *Nat Genet*, 35 (1), 25-31.
Osheroff, WP, et al. (1999), 'The fidelity of DNA polymerase beta during distributive and processive DNA synthesis.', *J Biol Chem*, 274 (6), 3642-50.
Otwinowski, Z, et al. (2003), 'Multiparametric scaling of diffraction intensities.', *Acta Crystallogr A*, 59 (Pt 3), 228-34.
Pages, V and RP Fuchs (2002), 'How DNA lesions are turned into mutations within cells?', *Oncogene*, 21 (58), 8957-66.
Papavasiliou, F, et al. (1997), 'V(D)J recombination in mature B cells: a mechanism for altering antibody responses.', *Science*, 278 (5336), 298-301.
Paques, F and JE Haber (1999), 'Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae.', *Microbiol Mol Biol Rev*, 63 (2), 349-404.
Parsons, JL, et al. (2008), 'CHIP-mediated degradation and DNA damage-dependent stabilization regulate base excision repair proteins.', *Mol Cell*, 29 (4), 477-87.
Parsons, JL, et al. (2009), 'Ubiquitin ligase ARF-BP1/Mule modulates base excision repair.', *EMBO J*, 28 (20), 3207-15.
Patel, SS, I Wong, and KA Johnson (1991), 'Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant.', *Biochemistry*, 30 (2), 511-25.
Pavlov, YI, PV Shcherbakova, and IB Rogozin (2006), 'Roles of DNA polymerases in replication, repair, and recombination in eukaryotes.', *Int Rev Cytol*, 255 41-132.
Peitsch, MC, et al. (1995), 'The Swiss-3DImage collection and PDB-Browser on the World-Wide Web.', *Trends Biochem Sci*, 20 (2), 82-84.
Pelletier, H, et al. (1994), 'Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP.', *Science*, 264 (5167), 1891-903.
Pelletier, H and MR Sawaya (1996), 'Characterization of the metal ion binding helix-hairpin-helix motifs in human DNA polymerase beta by X-ray structural analysis.', *Biochemistry*, 35 (39), 12778-87.
Pelletier, H, et al. (1996), 'A structural basis for metal ion mutagenicity and nucleotide selectivity in human DNA polymerase beta.', *Biochemistry*, 35 (39), 12762-77.
Pelletier, H, et al. (1996), 'Crystal structures of human DNA polymerase beta complexed with DNA: implications for catalytic mechanism, processivity, and fidelity.', *Biochemistry*, 35 (39), 12742-61.
Pfeiffer, P, et al. (1994), 'Resolution and conservation of mismatches in DNA end joining.', *Mutagenesis*, 9 (6), 527-35.
Picher, AJ, et al. (2006), 'Promiscuous mismatch extension by human DNA polymerase lambda.', *Nucleic Acids Res*, 34 (11), 3259-66.
Picher, AJ and L Blanco (2007), 'Human DNA polymerase lambda is a proficient extender of primer ends paired to 7,8-dihydro-8-oxoguanine.', *DNA Repair (Amst)*, 6 (12), 1749-56.
Pitcher, RS, et al. (2005), 'Domain structure of a NHEJ DNA repair ligase from Mycobacterium tuberculosis.', *J Mol Biol*, 351 (3), 531-44.
Pitcher, RS, TE Wilson, and AJ Doherty (2005), 'New insights into NHEJ repair processes in prokaryotes.', *Cell Cycle*, 4 (5), 675-78.
Pitcher, RS, et al. (2006), 'Mycobacteriophage exploit NHEJ to facilitate genome circularization.', *Mol Cell*, 23 (5), 743-48.
Pitcher, RS, NC Brissett, and AJ Doherty (2007), 'Nonhomologous end-joining in bacteria: a microbial perspective.', *Annu Rev Microbiol*, 61 259-82.
Pitcher, RS, et al. (2007), 'Structure and function of a mycobacterial NHEJ DNA repair polymerase.', *J Mol Biol*, 366 (2), 391-405.
Pitcher, RS, et al. (2007), 'NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation.', *DNA Repair (Amst)*, 6 (9), 1271-76.
Prakash, S, RE Johnson, and L Prakash (2005), 'Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function.', *Annu Rev Biochem*, 74 317-53.
Prasad, R, WA Beard, and SH Wilson (1994), 'Studies of gapped DNA substrate binding by mammalian DNA polymerase beta. Dependence on 5'-phosphate group.', *J Biol Chem*, 269 (27), 18096-101.
Prasad, R, et al. (1996), 'Specific interaction of DNA polymerase beta and DNA ligase I in a multiprotein base excision repair complex from bovine testis.', *J Biol Chem*, 271 (27), 16000-07.
Prasad, R, et al. (1998), 'Human DNA polymerase beta deoxyribose phosphate lyase. Substrate specificity and catalytic mechanism.', *J Biol Chem*, 273 (24), 15263-70.
Prasad, R, et al. (1998), 'Functional analysis of the amino-terminal 8-kDa domain of DNA polymerase beta as revealed by site-directed mutagenesis. DNA binding and 5'-deoxyribose phosphate lyase activities.', *J Biol Chem*, 273 (18), 11121-26.
Prasad, R, et al. (2005), 'Structural insight into the DNA polymerase beta deoxyribose phosphate lyase mechanism.', *DNA Repair (Amst)*, 4 (12), 1347-57.
Radford, SJ, et al. (2007), 'Heteroduplex DNA in meiotic recombination in Drosophila mei-9 mutants.', *Genetics*, 176 (1), 63-72.

Ramsden, DA (2011), 'Polymerases in nonhomologous end joining: building a bridge over broken chromosomes.', *Antioxid Redox Signal*, 14 (12), 2509-19.
Ranganathan, V, et al. (2001), 'Rescue of a telomere length defect of Nijmegen breakage syndrome cells requires NBS and telomerase catalytic subunit.', *Curr Biol*, 11 (12), 962-66.
Rechkoblit, O, et al. (2006), 'Stepwise translocation of Dpo4 polymerase during error-free bypass of an oxoG lesion.', *PLoS Biol*, 4 (1), e11.
Riha, K, ML Heacock, and DE Shippen (2006), 'The role of the nonhomologous end-joining DNA double-strand break repair pathway in telomere biology.', *Annu Rev Genet*, 40 237-77.
Roepstorff, P and J Fohlman (1984), 'Proposal for a common nomenclature for sequence ions in mass spectra of peptides.', *Biomed Mass Spectrom*, 11(11) (11), 601.
Roettger, MP, et al. (2004), 'Pre-steady-state kinetic studies of the fidelity of human DNA polymerase mu.', *Biochemistry*, 43 (43), 13827-38.
Romain, F, et al. (2009), 'Conferring a template-dependent polymerase activity to terminal deoxynucleotidyltransferase by mutations in the Loop1 region.', *Nucleic Acids Res*, 37 (14), 4642-56.
Rooney, S, et al. (2002), 'Leaky Scid phenotype associated with defective V(D)J coding end processing in Artemis-deficient mice.', *Mol Cell*, 10 (6), 1379-90.
Roth, DB and JH Wilson (1986), 'Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction.', *Mol Cell Biol*, 6 (12), 4295-304.
Rouet, P, F Smih, and M Jasin (1994), 'Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease.', *Mol Cell Biol*, 14 (12), 8096-106.
Roychoudhury, R (1972), 'Enzymic synthesis of polynucleotides. Oligodeoxynucleotides with one 3'-terminal ribonucleotide as primers for polydeoxynucleotide synthesis.', *J Biol Chem*, 247 (12), 3910-17.
Ruiz, JF, et al. (2001), 'DNA polymerase mu, a candidate hypermutase?', *Philos Trans R Soc Lond B Biol Sci*, 356 (1405), 99-109.
Ruiz, JF, et al. (2003), 'Lack of sugar discrimination by human Pol mu requires a single glycine residue.', *Nucleic Acids Res*, 31 (15), 4441-49.
Ruiz, JF, et al. (2004), 'Overexpression of human DNA polymerase mu (Pol mu) in a Burkitt's lymphoma cell line affects the somatic hypermutation rate.', *Nucleic Acids Res*, 32 (19), 5861-73.
Salas, M (1991), 'Protein-priming of DNA replication.', *Annu Rev Biochem*, 60 39-71.
Sawaya, MR, et al. (1994), 'Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism.', *Science*, 264 (5167), 1930-35.
Sawaya, MR, et al. (1997), 'Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism.', *Biochemistry*, 36 (37), 11205-15.
Sawchuk, DJ, et al. (1997), 'V(D)J recombination: modulation of RAG1 and RAG2 cleavage activity on 12/23 substrates by whole cell extract and DNA-bending proteins.', *J Exp Med*, 185 (11), 2025-32.
Sawyer, SL and HS Malik (2006), 'Positive selection of yeast nonhomologous end-joining genes and a retrotransposon conflict hypothesis.', *Proc Natl Acad Sci U S A*, 103 (47), 17614-19.
Schar, P, et al. (1997), 'A newly identified DNA ligase of Saccharomyces cerevisiae involved in RAD52-independent repair of DNA double-strand breaks.', *Genes Dev*, 11 (15), 1912-24.
Schatz, DG, MA Oettinger, and D Baltimore (1989), 'The V(D)J recombination activating gene, RAG-1.', *Cell*, 59 (6), 1035-48.
Schofield, MJ and P Hsieh (2003), 'DNA mismatch repair: molecular mechanisms and biological function.', *Annu Rev Microbiol*, 57 579-608.
Seki, M, F Marini, and RD Wood (2003), 'POLQ (Pol theta), a DNA polymerase and DNA-dependent ATPase in human cells.', *Nucleic Acids Res*, 31 (21), 6117-26.
Semizarov, DG, et al. (1997), 'Stereoisomers of deoxynucleoside 5'-triphosphates as substrates for template-dependent and -independent DNA polymerases.', *J Biol Chem*, 272 (14), 9556-60.
Sharief, FS, et al. (1999), 'Cloning and chromosomal mapping of the human DNA polymerase theta (POLQ), the eighth human DNA polymerase.', *Genomics*, 59 (1), 90-96.
Shevchenko, A, et al. (1996), 'A strategy for identifying gel-separated proteins in sequence databases by MS alone.', *Biochem Soc Trans*, 24 (3), 893-96.
Shevelev, I, et al. (2003), 'Mutagenesis of human DNA polymerase lambda: essential roles of Tyr505 and Phe506 for both DNA polymerase and terminal transferase activities.', *Nucleic Acids Res*, 31 (23), 6916-25.
Shevelev, IV and U Hubscher (2002), 'The 3' 5' exonucleases.', *Nat Rev Mol Cell Biol*, 3 (5), 364-76.
Shinkai, Y, et al. (1992), 'RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement.', *Cell*, 68 (5), 855-67.
Shinohara, A and T Ogawa (1995), 'Homologous recombination and the roles of double-strand breaks.', *Trends Biochem Sci*, 20 (10), 387-91.
Showalter, AK, et al. (2001), 'Solution structure of a viral DNA polymerase X and evidence for a mutagenic function.', *Nat Struct Biol*, 8 (11), 942-46.
Shrivastav, M, LP De Haro, and JA Nickoloff (2008), 'Regulation of DNA double-strand break repair pathway choice.', *Cell Res*, 18 (1), 134-47.

Silvian, LF, et al. (2001), 'Crystal structure of a DinB family error-prone DNA polymerase from Sulfolobus solfataricus.', *Nat Struct Biol*, 8 (11), 984-89.
Singhal, RK and SH Wilson (1993), 'Short gap-filling synthesis by DNA polymerase beta is processive.', *J Biol Chem*, 268 (21), 15906-11.
Smogorzewska, A, et al. (2002), 'DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2.', *Curr Biol*, 12 (19), 1635-44.
Sobol, RW, et al. (1996), 'Requirement of mammalian DNA polymerase-beta in base-excision repair.', *Nature*, 379 (6561), 183-86.
Sobol, RW, et al. (2000), 'The lyase activity of the DNA repair protein beta-polymerase protects from DNA-damage-induced cytotoxicity.', *Nature*, 405 (6788), 807-10.
Sobol, RW and SH Wilson (2001), 'Mammalian DNA beta-polymerase in base excision repair of alkylation damage.', *Prog Nucleic Acid Res Mol Biol*, 68 57-74.
Sousa, R, et al. (1993), 'Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution.', *Nature*, 364 (6438), 593-99.
Steitz, TA and JA Steitz (1993), 'A general two-metal-ion mechanism for catalytic RNA.', *Proc Natl Acad Sci U S A*, 90 (14), 6498-502.
Steitz, TA, et al. (1994), 'A unified polymerase mechanism for nonhomologous DNA and RNA polymerases.', *Science*, 266 (5193), 2022-25.
Stracker, TH, CT Carson, and MD Weitzman (2002), 'Adenovirus oncoproteins inactivate the Mre11-Rad50-NBS1 DNA repair complex.', *Nature*, 418 (6895), 348-52.
Streisinger, G, et al. (1966), 'Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday.', *Cold Spring Harb Symp Quant Biol*, 31 77-84.
Sugo, N, et al. (2000), 'Neonatal lethality with abnormal neurogenesis in mice deficient in DNA polymerase beta.', *EMBO J*, 19 (6), 1397-404.
Sukumar, N, et al. (2000), 'Crystallization of the catalytic domain of murine terminal deoxynucleotidyl transferase.', *Acta Crystallogr D Biol Crystallogr*, 56 (Pt 12), 1662-64.
Symington, LS (2002), 'Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair.', *Microbiol Mol Biol Rev*, 66 (4), 630-70, table of contents.
Takeuchi, R, et al. (2006), 'Drosophila DNA polymerase zeta interacts with recombination repair protein 1, the Drosophila homologue of human abasic endonuclease 1.', *J Biol Chem*, 281 (17), 11577-85.
Taladriz, S, et al. (2001), 'Nuclear DNA polymerase beta from Leishmania infantum. Cloning, molecular analysis and developmental regulation.', *Nucleic Acids Res*, 29 (18), 3822-34.
Tanabe, K, EW Bohn, and SH Wilson (1979), 'Steady-state kinetics of mouse DNA polymerase beta.', *Biochemistry*, 18 (15), 3401-06.
Teo, SH and SP Jackson (1997), 'Identification of Saccharomyces cerevisiae DNA ligase IV: involvement in DNA double-strand break repair.', *EMBO J*, 16 (15), 4788-95.
Thode, S, et al. (1990), 'A novel pathway of DNA end-to-end joining.', *Cell*, 60 (6), 921-28.
Thompson, EH, et al. (2002), 'Determinants of DNA mismatch recognition within the polymerase domain of the Klenow fragment.', *Biochemistry*, 41 (3), 713-22.
Tippin, B, et al. (2004), 'To slip or skip, visualizing frameshift mutation dynamics for error-prone DNA polymerases.', *J Biol Chem*, 279 (44), 45360-68.
Tonegawa, S (1983), 'Somatic generation of antibody diversity.', *Nature*, 302 (5909), 575-81.
Traut, TW (1994), 'Physiological concentrations of purines and pyrimidines.', *Mol Cell Biochem*, 140 (1), 1-22.
Tseng, HM and AE Tomkinson (2002), 'A physical and functional interaction between yeast Pol4 and Dnl4-Lif1 links DNA synthesis and ligation in nonhomologous end joining.', *J Biol Chem*, 277 (47), 45630-37.
Uchiyama, Y, Y Suzuki, and K Sakaguchi (2008), 'Characterization of plant XRCC1 and its interaction with proliferating cell nuclear antigen.', *Planta*, 227 (6), 1233-41.
Uchiyama, Y, et al. (2009), 'Distribution and roles of X-family DNA polymerases in eukaryotes.', *Biochimie*, 91 (2), 165-70.
Uljon, SN, et al. (2004), 'Crystal structure of the catalytic core of human DNA polymerase kappa.', *Structure*, 12 (8), 1395-404.
Ulrich, HD (2004), 'How to activate a damage-tolerant polymerase: consequences of PCNA modifications by ubiquitin and SUMO.', *Cell Cycle*, 3 (1), 15-18.
Unoki, M and Y Nakamura (2001), 'Growth-suppressive effects of BPOZ and EGR2, two genes involved in the PTEN signaling pathway.', *Oncogene*, 20 (33), 4457-65.
van Gent, DC, et al. (1997), 'Stimulation of V(D)J cleavage by high mobility group proteins.', *EMBO J*, 16 (10), 2665-70.
van Loon, B and U Hubscher (2009), 'An 8-oxo-guanine repair pathway coordinated by MUTYH glycosylase and DNA polymerase lambda.', *Proc Natl Acad Sci U S A*, 106 (43), 18201-06.
Walker, JR, RA Corpina, and J Goldberg (2001), 'Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair.', *Nature*, 412 (6847), 607-14.
Ward, JF (1981), 'Some biochemical consequences of the spatial distribution of ionizing radiation-produced free radicals.', *Radiat Res*, 86 (2), 185-95.
Ward, JF (2000), 'Complexity of damage produced by ionizing radiation.', *Cold Spring Harb Symp Quant Biol*, 65 377-82.
Watson, JD and FH Crick (1953), 'The structure of DNA.', *Cold Spring Harb Symp Quant Biol*, 18 123-31.

Weitzman, MD, et al. (2004), 'Interactions of viruses with the cellular DNA repair machinery.', *DNA Repair (Amst)*, 3 (8-9), 1165-73.
Weller, GR and AJ Doherty (2001), 'A family of DNA repair ligases in bacteria?', *FEBS Lett*, 505 (2), 340-42.
Weller, GR, et al. (2002), 'Identification of a DNA nonhomologous end-joining complex in bacteria.', *Science*, 297 (5587), 1686-89.
Werneburg, BG, et al. (1996), 'DNA polymerase beta: pre-steady-state kinetic analysis and roles of arginine-283 in catalysis and fidelity.', *Biochemistry*, 35 (22), 7041-50.
West, SC (2003), 'Molecular views of recombination proteins and their control.', *Nat Rev Mol Cell Biol*, 4 (6), 435-45.
Wilson, DM 3rd and DR McNeill (2007), 'Base excision repair and the central nervous system.', *Neuroscience*, 145 (4), 1187-200.
Wilson, RC and JD Pata (2008), 'Structural insights into the generation of single-base deletions by the Y family DNA polymerase dbh.', *Mol Cell*, 29 (6), 767-79.
Wilson, SH and TA Kunkel (2000), 'Passing the baton in base excision repair.', *Nat Struct Biol*, 7 (3), 176-78.
Wilson, TE, U Grawunder, and MR Lieber (1997), 'Yeast DNA ligase IV mediates non-homologous DNA end joining.', *Nature*, 388 (6641), 495-98.
Wilson, TE and MR Lieber (1999), 'Efficient processing of DNA ends during yeast nonhomologous end joining. Evidence for a DNA polymerase beta (Pol4)-dependent pathway.', *J Biol Chem*, 274 (33), 23599-609.
Wimmer, U, et al. (2008), 'Control of DNA polymerase lambda stability by phosphorylation and ubiquitination during the cell cycle.', *EMBO Rep*, 9 (10), 1027-33.
Wong, I and TM Lohman (1993), 'A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions.', *Proc Natl Acad Sci U S A*, 90 (12), 5428-32.
Yakovleva, L and S Shuman (2006), 'Nucleotide misincorporation, 3'-mismatch extension, and responses to abasic sites and DNA adducts by the polymerase component of bacterial DNA ligase D.', *J Biol Chem*, 281 (35), 25026-40.
Yan, CT, et al. (2007), 'IgH class switching and translocations use a robust non-classical end-joining pathway.', *Nature*, 449 (7161), 478-82.
Yaneva, M, T Kowalewski, and MR Lieber (1997), 'Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies.', *EMBO J*, 16 (16), 5098-112.
Yang, W (2003), 'Damage repair DNA polymerases Y.', *Curr Opin Struct Biol*, 13 (1), 23-30.
Yang, W (2005), 'Portraits of a Y-family DNA polymerase.', *FEBS Lett*, 579 (4), 868-72.
Yang, W and R Woodgate (2007), 'What a difference a decade makes: insights into translesion DNA synthesis.', *Proc Natl Acad Sci U S A*, 104 (40), 15591-98.
Yu, X, et al. (2003), 'The BRCT domain is a phospho-protein binding domain.', *Science*, 302 (5645), 639-42.
Zhang, H, et al. (2007), 'Fluorescence of 2-aminopurine reveals rapid conformational changes in the RB69 DNA polymerase-primer/template complexes upon binding and incorporation of matched deoxynucleoside triphosphates.', *Nucleic Acids Res*, 35 (18), 6052-62.
Zhang, Y, et al. (2001), 'Highly frequent frameshift DNA synthesis by human DNA polymerase mu.', *Mol Cell Biol*, 21 (23), 7995-8006.
Zhang, Y, et al. (2002), 'Lesion bypass activities of human DNA polymerase mu.', *J Biol Chem*, 277 (46), 44582-87.
Zhou, BL, JD Pata, and TA Steitz (2001), 'Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain.', *Mol Cell*, 8 (2), 427-37.
Zhou, RZ, et al. (2008), 'Tolerance for 8-oxoguanine but not thymine glycol in alignment-based gap filling of partially complementary double-strand break ends by DNA polymerase lambda in human nuclear extracts.', *Nucleic Acids Res*, 36 (9), 2895-905.
Zhu, C, et al. (1996), 'Ku86-deficient mice exhibit severe combined immunodeficiency and defective processing of V(D)J recombination intermediates.', *Cell*, 86 (3), 379-89.
Zhu, H, et al. (2006), 'Atomic structure and nonhomologous end-joining function of the polymerase component of bacterial DNA ligase D.', *Proc Natl Acad Sci U S A*, 103 (6), 1711-16.



# APPENDIX: PUBLICATIONS

# Limited terminal transferase in human DNA polymerase $\mu$ defines the required balance between accuracy and efficiency in NHEJ

**Paula Andrade, María José Martín, Raquel Juárez, Francisco López de Saro, and Luis Blanco[1]**

Centro de Biología Molecular Severo Ochoa, Consejo Superior de Investigaciones Científicas-Universidad Autónoma de Madrid, 28049 Madrid, Spain

Communicated by Margarita Salas, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain, July 30, 2009 (received for review May 25, 2009)

**DNA polymerase mu (Pol$\mu$) is a family X member implicated in DNA repair, with template-directed and terminal transferase (template-independent) activities. It has been proposed that the terminal transferase activity of Pol$\mu$ can be specifically required during non-homologous end joining (NHEJ) to create or increase complementarity of DNA ends. By site-directed mutagenesis in human Pol$\mu$, we have identified a specific DNA ligand residue (Arg$^{387}$) that is responsible for its limited terminal transferase activity compared to that of human TdT, its closest homologue (42% amino acid identity). Pol$\mu$ mutant R387K (mimicking TdT) displayed a large increase in terminal transferase activity, but a weakened interaction with ssDNA. That paradox can be explained by the regulatory role of Arg$^{387}$ in the translocation of the primer from a non-productive E:DNA complex to a productive E:DNA:dNTP complex in the absence of a templating base, which is probably the rate limiting step during template-independent synthesis. Further, we show that the Pol$\mu$ switch from terminal transferase to templated insertions in NHEJ reactions is triggered by recognition of a 5′-P at a second DNA end, whose 3′-protrusion could provide a templating base to facilitate such a special "pre-catalytic translocation step." These studies shed light on the mechanism by which a rate-limited terminal transferase activity in Pol$\mu$ could regulate the balance between accuracy and necessary efficiency, providing some variability during NHEJ.**

Physical and chemical damage of DNA is the root cause of a large number of human syndromes, including premature aging, various cancer predispositions, and congenital abnormalities. To overcome DNA damage and maintain stability in the chromosomes, cells contain an array of specific DNA repair pathways that act upon the different kinds of lesions (1). Specialized DNA polymerases are essential actors within these pathways and, in humans, at least 12 are devoted to overcome or repair DNA damage in the cell (2). DNA polymerases of the X family, which in mammals include DNA polymerases beta (Polß), lambda (Pol$\lambda$), mu (Pol$\mu$), and terminal deoxynucleotidyl transferase (TdT), are structurally related enzymes specialized in repair pathways involving double strand breaks (DSB) and gaps (3).

Unlike Polß, Pol$\lambda$, Pol$\mu$, and TdT contain a BRCA-1 C-terminal (BRCT) interaction domain at their N-termini (4–7), that serves to interact with the core NHEJ factors Ku and XRCC4-Ligase IV (8–12), suggesting their involvement in DSB repair via NHEJ. Whereas DSB repair by homologous recombination is highly accurate because it relies directly on complementary DNA sequences, the accuracy of the repair reaction by NHEJ is lower, somehow compromising preservation of the original DNA information. In this context, NHEJ DNA polymerases would be needed either to extend the 3′-end at 5′-overhangs, to fill gaps created by limited complementarity of the two DNA ends, or eventually in the case of joining non-complementary (incompatible) ends, to add sequences at the 3′-end to create a minimal cohesive base pairing (13). To accomplish these tasks, Pol$\lambda$, Pol$\mu$, and TdT have a gradient of template dependency (12): Pol$\lambda$ requires at least one complementary base pair between the two DNA ends (microhomology) to be able to fill gaps as template-directed insertion events; Pol$\mu$ can realign DNA strands allowing distortions to maximize cohesion (14), and can even extend a 3′-end terminus with minimal or null complementarity, contributing to the efficiency of end-joining with either templated or untemplated insertions (15); TdT, on the other hand, is a strong template-independent polymerase, displaying only a strict terminal transferase activity. TdT function is essential during V(D)J recombination, a specialized variant of the NHEJ pathway that takes place during the maturation of antibodies in B lymphocytes, in which the creation of diversity (by addition of random DNA sequences) is favoured over accuracy in the repair of the DSB (16).

Pol$\mu$ is the only DNA polymerase known that combines template-dependent and template-independent (terminal transferase) activities in one enzyme (6). The relative importance of the terminal transferase activity of Pol$\mu$ during NHEJ has been controversial and probably underscored as it is difficult to distinguish template-dependent from random additions, if the latter are successful in creating complementarity between the two ends of a DSB (12). However, there are insightful structure-function correlations suggesting that terminal transferase activity is important for NHEJ. Thus, *loop1* in human Pol$\mu$, likely more flexible than that present in TdT (17), was shown to be important for both NHEJ (12) and terminal transferase activity (13). As early proposed by Juárez et al. (13), Pol$\mu$ can alternate between stable interactions with ssDNA vs. dsDNA via conformational changes in its flexible *loop1*, being disordered in the crystal structure of Pol$\mu$ in complex with gap DNA (18). This structure of a ternary complex of human Pol$\mu$ allowed proposing that a specific histidine residue (His$^{329}$ in Pol$\mu$ and His$^{342}$ in TdT; absent in Polß and Pol$\lambda$) plays an important role in terminal transferase activity (in both Pol$\mu$ and TdT) and NHEJ of incompatible ends mediated by Pol$\mu$ (18).

In this study, we identified a specific arginine in human Pol$\mu$ (Arg$^{387}$) that is responsible for its limited terminal transferase activity in comparison with that of TdT. One of the mutations studied (R387K), which greatly increased terminal transferase activity in Pol$\mu$, and at the same time reduced interaction with the primer strand, gives insights into the mechanism and details of the reaction. We propose that movement of the primer from a non-productive E:DNA complex to a productive E:DNA:dNTP complex is probably a rate limiting step for the terminal transferase addition of nucleotides. The mild terminal transferase activity of Pol$\mu$ contributes to establish a minimal

Author contributions: R.J., F.L.d.S., and L.B. designed research; P.A., M.J.M., and R.J. performed research; R.J., F.L.d.S., and L.B. analyzed data; and L.B. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: lblanco@cbm.uam.es.

This article contains supporting information online at www.pnas.org/cgi/content/full/0908492106/DCSupplemental.

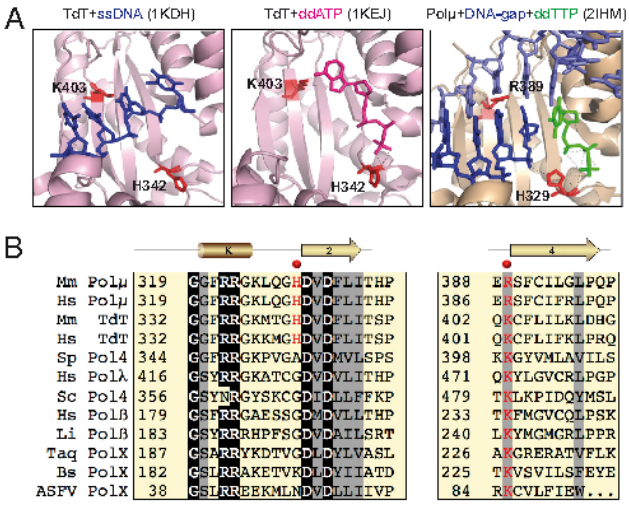



**Fig. 1.** Specific residues involved in terminal transferase. (*A*) Crystal structures of TdT, either complexed with ssDNA (1KDH) or dNTP (1KEJ), and Pol$\mu$ (1IHM). Residues selected for mutagenesis in Pol$\mu$, and their orthologues in TdT, are indicated in red sticks. DNA substrates are indicated in dark (primer strand) and light (template and downstream strand) blue. Incoming ddNTPs are indicated in magenta (non templated) or green (templated). (*B*) Amino acid sequence alignment of family X polymerases from different species (Hs, human; Mm, mouse; Sp, *Schizosaccharomyces pombe*; Sc, *Saccharomyces cerevisiae*; Li, *Leishmania infantum*; Taq, *Thermus aquaticus*; Bs, *Bacillus subtilis*; and ASFV, african swine fever virus) along the indicated regions. Residues selected for mutagenesis in Pol$\mu$ are indicated with red dots. Invariant residues are depicted in white over black background, and conservative residues have gray background.

cohesion of DNA ends but limiting excessive mutagenic events during NHEJ.

## Results

**Search for Specific Residues Involved in Terminal Transferase Versus Template-Directed Polymerization.** Pol$\mu$ and TdT share 42% overall amino acid sequence identity and are highly conserved in vertebrates (6). Accordingly, the crystallographic structure of murine Pol$\mu$ and TdT is very similar, even in very different contexts (see Fig. 1*A*): either bound to a primer (1KDH) or to a nucleotide (1KEJ), as in the case of TdT (17); or on a gapped DNA substrate with an incoming nucleotide (2IHM), as in the murine Pol$\mu$ crystal (18). However, a more detailed inspection of the different TdT and Pol$\mu$ structures, and their comparison with those of human Pol$\lambda$ and Pol$\beta$, pointed out to two specific residues in murine Pol$\mu$, His$^{329}$, and Arg$^{389}$, that could display dual interactions, thus having a specific potential in regulating terminal transferase versus template-directed nucleotide additions. As indicated in Fig. 1*A*, that histidine in TdT (His$^{342}$) is not involved in a direct interaction with the primer (left panel), but interacts with the incoming nucleotide (central panel), and its substitution abolishes terminal transferase activity (18). However, murine Pol$\mu$ His$^{329}$ is hydrogen-bonded to both primer terminus and nucleotide (right panel), and its proposed function is to align them properly before catalysis (18). As shown in Fig. 1*B*, no equivalent histidine is present in either Pol$\beta$ or Pol$\lambda$, or in any other PolX enzyme.

Murine Pol$\mu$ Arg$^{389}$ is interacting with the template strand, contacting the sugar at residue in position $-3$ with respect to the templating base. An identical interaction of the equivalent residue in Pol$\lambda$ (Lys$^{472}$) and Pol$\beta$ (Lys$^{234}$) is observed in ternary complexes of these enzymes (1XSN;1BPY). As shown in Fig. 1*B*, an arginine residue is unique to Pol$\mu$, whereas all other family

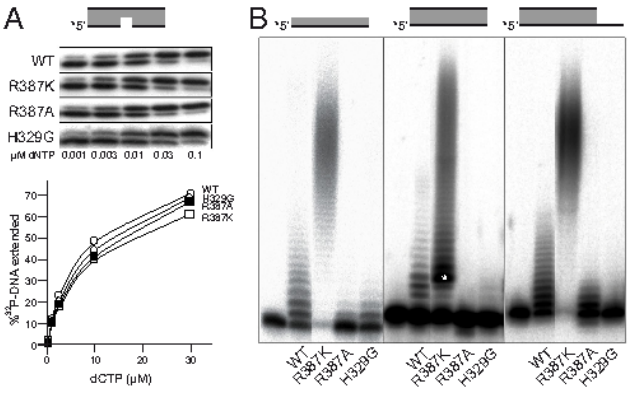



**Fig. 2.** Arg$^{387}$ regulates the rate of terminal transferase activity in human Pol$\mu$, while His$^{329}$ is essential. (*A*) Polymerization reactions on a gapped-DNA substrate (see scheme) was carried out in the presence of 2.5 mM $MgCl_2$, the indicated amounts of dCTP, and 300 nM of either Pol$\mu$ wt or the indicated mutants. After incubation for 30 min at 30 °C, extension of the 5′P-labeled primer (indicated with an asterisk) was analyzed by 8 M urea-20% PAGE and autoradiography, and the results quantitated by densitometry. (*B*) Terminal transferase activity on different substrates (ssDNA/polyT, blunt dsDNA and 3′-protruding (9T) dsDNA; 5 nM) was assayed for 30 min at 30 °C in the presence of 1 mM $MnCl_2$, 100 $\mu$M dTTP, and 600 nM of either Pol$\mu$ wt or the indicated mutants, and analyzed by 8 M urea-20% PAGE and autoradiography.

members have a lysine instead. Interestingly, that lysine in TdT (Lys$^{403}$) is a direct ligand of the primer strand, as seen in Fig. 1*A* (left panel). No experimental evidence for the functional importance of this positively charged residue (either arginine or lysine) was yet available. Therefore, the equivalent residues in human Pol$\mu$, His$^{329}$, and Arg$^{387}$, were targeted for site-directed mutagenesis. His$^{329}$ was changed to glycine, as it naturally occurs in human Pol$\beta$ and Pol$\lambda$, whereas Arg$^{387}$ was changed either to alanine (R387A), to eliminate the positive charge, or to lysine (R387K), as it occurs in all other DNA polymerases from family X.

**Arg$^{387}$ Regulates the Rate of Terminal Transferase Activity in Human Pol$\mu$, While His$^{329}$ Is Essential.** The selected mutants, obtained and purified as indicated in *Materials and Methods*, were tested for a differential effect on their template-dependent versus template-independent polymerization capacity. None of the mutants tested showed a significant defect in gap-filling (Fig. 2*A*), suggesting that residues His$^{329}$ and Arg$^{387}$ do not play a critical role in template-directed synthesis. Conversely, when terminal transferase was measured on various substrates, the effect of the mutations was dramatic (Fig. 2*B*). Mutant H329G was largely affected on any substrate compared to wild-type enzyme, in support of its specific relevance for terminal transferase reactions, as previously shown by using a different mutation [H329A; (18)]. On the contrary, in the case of Arg$^{387}$ the two substitutions made had opposite effects: whereas mutant R387A largely decreased terminal transferase activity, mutant R387K (mimicking TdT) produced a very significant increase of this template-independent reaction (Fig. 2*B*). These results suggest that the specific arginine that is only present in Pol$\mu$ could act as a regulator, reducing the catalytic efficiency of its intrinsic terminal transferase. Analysis of the kinetic parameters for mutant R387K (Table 1) showed a 38-fold increase in catalytic efficiency for inserting dTTP, not due to a difference in $K_m$, but to a large improvement in kcat. A similar conclusion was obtained by using the other three different dNTPs, provided independently (Fig. S1 and Table 1). The above results indicate that the histidine residue has an essential role during terminal transferase addition in both TdT (His$^{342}$) and Pol$\mu$ (His$^{329}$), whereas the specific arginine in Pol$\mu$ (Arg$^{387}$) somehow limits the catalytic step of the

**Table 1. Kinetic parameters of the terminal transferase activity of mutant R387K with any of the four dNTPs**

| | | $Km_{app}$, μM | $K_{cat}$, $s^{-1}$ | Catalytic efficiency | $f^0$ ext |
|---|---|---|---|---|---|
| Polμ WT | dTTP | 319 ± 16 | 0.0004 ± 0.0002 | (1.39 ± 0.0004) $10^{-6}$ | 1 |
| | dATP | 4,300 ± 2400 | 0.0016 ± 0.0006 | (4.10 ± 0.92) $10^{-7}$ | 1 |
| | dCTP | 224 ± 26 | 0.0008 ± 0.0004 | (5.97 ± 2.35) $10^{-6}$ | 1 |
| | dGTP | 5,830 ± 70 | 0.0049 ± 0.0025 | (8.43 ± 4.34) $10^{-7}$ | 1 |
| R387K | dTTP | 307 ± 17 | 0.015 ± 0.006 | (5.35 ± 1.13) $10^{-5}$ | 38 |
| | dATP | 5,100 ± 2300 | 0.015 ± 0.00008 | (3.28 ± 1.47) $10^{-6}$ | 8 |
| | dCTP | 131 ± 24 | 0.018 ± 0.004 | (1.36 ± 0.34) $10^{-4}$ | 22 |
| | dGTP | 3,010 ± 100 | 0.188 ± 0.015 | (6.25 ± 0.30) $10^{-5}$ | 74 |

Terminal transferase activity was assayed either on polydT (with dTTP, dCTP and dGTP), or on polydA (with dATP), as described in *Materials and Methods*. Mean values (average) and standard deviation, corresponding to various independent experiments, are indicated.

terminal transferase reaction, perhaps explaining its low activity in comparison to TdT.

**Rate-Limiting Step for the Terminal Transferase: Translocation of the Primer before Catalysis.** By using nitrocellulose filter-binding assays (19) we assessed that human Polμ stably interacts with ssDNA (Fig. 3*A*), in a perhaps non-productive complex as that observed in the crystal structure of TdT (17). Paradoxically, mutant R387K, although having an increased terminal transferase activity, displayed an about 4-fold lower interaction with ssDNA (Fig. 3*A*); this defective binding was also observed for mutant R387A, but not in mutant H329G (Fig. 3*A*; wild-type Polμ Kdis = 1.3 μM; Polμ R387K Kdis = 5.5 μM; Polμ R387A Kdis = 5.8 μM; Polμ H329G Kdis = 1.3 μM).

The first clue of a plausible rate-limiting step for the terminal transferase derived from the crystal structure of TdT bound to ssDNA (1KDH), whose 3′-terminus was unexpectedly located in a postcatalytic situation, occupying the binding site for the next incoming dNTP (17; see Fig. 1*A*, left panel). No crystal structure of a precatalytic ternary complex for the terminal transferase reaction is available, neither for TdT nor for Polμ. Therefore, and combining the 3D information available for both TdT and Polμ, it can be proposed that the rate-limiting step of the terminal transferase most likely involves, for both enzymes, the movement of the single-stranded (3′-protruding) primer-terminus from a "non-productive" E:DNA complex to a "productive" E:DNA:dNTP complex (Fig. 3*B*), to achieve untemplated addition of nucleotides, irrespective of either a processive or a distributive cycling of the enzyme (see Fig. S2 for further details).

We next evaluated the stability of the ssDNA primer in the presence of metal and dNTP (that will allow formation of a productive complex), but impeding reaction by the presence of a ddNMP at the primer-terminus. Interestingly, binding of both wt and R387K mutant Polμ to ssDNA was weakened by the presence of dNTP (Fig. 3*B*), supporting the idea that there is a necessary movement of the primer (1 nt backwards) to form a productive complex, in which the dNTP has to compete and finally displace the 3′-terminus to occupy their corresponding binding sites, necessary for catalysis. Congruently, the intrinsic weaker binding to ssDNA of mutant R387K would facilitate primer movement associated to the rate limiting step, explaining the paradox of its high terminal transferase activity.

**Is Terminal Transferase Operating during NHEJ of Incompatible Ends? Importance of a Recessive 5′P.** It has been recently reported that Polμ's optimal end-joining of incompatible ends occurs at minimal (1–2 nt) protrusions; on longer protrusions, a single polymerase unit cannot maintain bridging of the two ends, and the yield of the reaction is lower, exclusively reflecting terminal transferase activity (15). Since recessive 5′ DNA ends must contain a 5′P group to allow

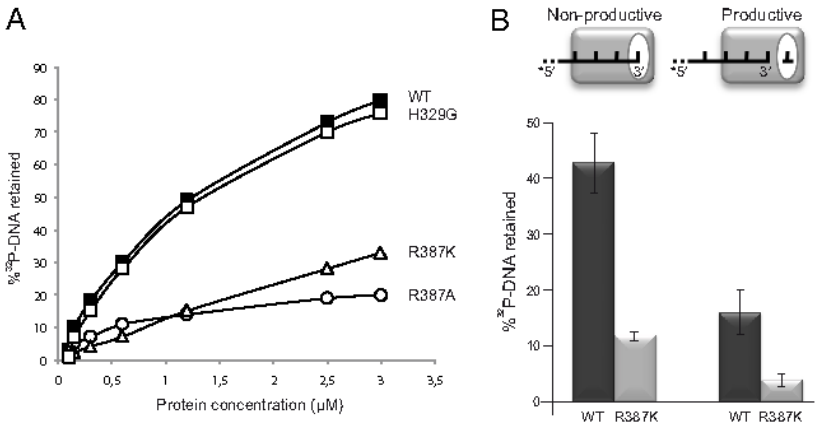



**Fig. 3.** Mutants at residue $Arg^{387}$ have a reduced ssDNA binding affinity. The presence of dNTP reduces ssDNA binding capacity by Polμ. (*A*) A 5′-labeled ssDNA (polydT; 5 nM) was used as substrate for interaction with different concentrations of either Polμ wt or mutants R387K, R387A, and H329G. After incubation for 10 min at 22 °C, the protein/ssDNA complexes were retained in nitrocellulose filters, as described in *Materials and Methods*. (*B*) Schemes depicting two different complexes: non-productive (nucleotide absent), indicative of a E:DNA binary complex in which the 3′ end of primer is occupying the nucleotide pocket (white oval); and productive (nucleotide present), indicative of a E:DNA:dNTP ternary complex in which the primer has been relocated. Binding capacity of Polμ wt and R387K mutant (600 nM in each case) to ssDNA/polyT, either in the absence (left) or presence (right) of dTTP, was measured by nitrocellulose filter binding assay, as described in *Materials and Methods*.

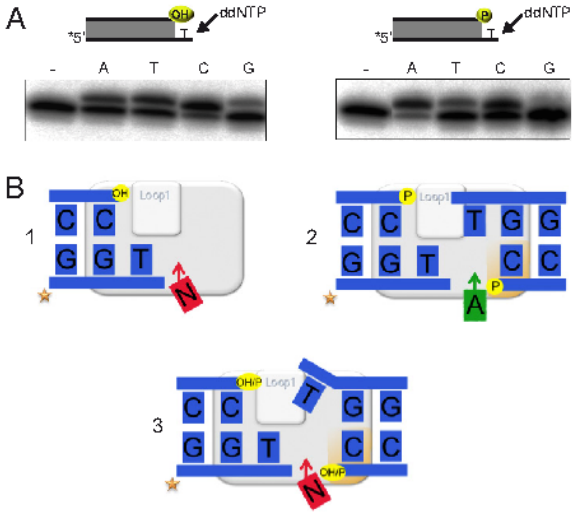



**Fig. 4.** Terminal transferase vs. NHEJ of incompatible ends. (*A*) Single nucleotide extension of a 3′-protruding (1T) dsDNA substrate (200 nM), either lacking (left panel) or having (right panel) a 5′-recessive P. The assay was performed as described in *Materials and Methods*, in the presence of 1 mM $MnCl_2$, 100 $\mu$M of the indicated ddNTP, and 200 nM of Pol$\mu$ wt. After incubation for 30 min at 30 °C, +1 extension of the 5′P-labeled oligonucleotide was analyzed by 8 M urea-20% PAGE and autoradiography. (*B*) The observed nucleotide specificity can be the consequence of either untemplated polymerization (terminal transferase; scheme 1) or DNA-directed additions (dA insertion templated in trans), thus achieving precise joining of two DNA ends (scheme 2). Scheme 3 shows a hybrid situation in which the joining reaction is untemplated, as the 3′-protruding template base provided in trans is not in a proper register. The gray rectangle represents a cartoon of Pol$\mu$, in which the specific loop 1, involved in both terminal transferase and connectivity during NHEJ of incompatible ends, is highlighted. The orange area in Pol$\mu$ represents the 8-kDa domain, specifically involved in 5′P recognition.

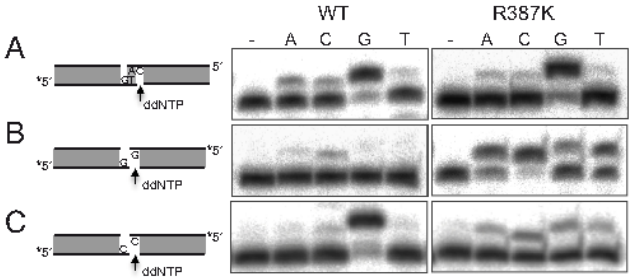



**Fig. 5.** R387K has a low fidelity during NHEJ of incompatible ends. (*A*) Accurate NHEJ of minimally complementary ends, using a 5′-labeled 3′-protruding (GT) dsDNA substrate (5 nM), and a cold 3′-protruding (CA) dsDNA substrate (25 nM), both having a recessive 5′P. Mutant R387K displayed an accurate behavior comparable to wild-type Pol$\mu$. (*B*) Inaccurate NHEJ of incompatible ends, using a 5′-labeled 3′-protruding (G) dsDNA substrate (5 nM), having a recessive 5′P. Mutant R387K was more efficient but displayed an inaccurate behavior comparable to wild-type Pol$\mu$. (*C*) Accurate vs. inaccurate NHEJ of incompatible ends, using a 5′-labeled 3′-protruding (C) dsDNA substrate (5 nM), having a recessive 5′P. Mutant R387K displayed an error-prone behavior compared to the accuracy of wild-type Pol$\mu$. In all cases, assays were performed in the presence of 2.5 mM $MgCl_2$, 100 $\mu$M of the indicated ddNTP, and 200 nM of either Pol$\mu$ wt or R387K mutant. After incubation for 30 min at 30 °C, +1 extension of the 5′P-labeled primer was analyzed by 8 M urea-20% PAGE and autoradiography.

ligation and repair of the DNA break, and family X polymerases contain an 8-kDa domain that interacts directly with the 5′P group of the downstream strand (3), we determined if the presence of a 5′P group modulates Pol$\mu$'s terminal transferase activity, perhaps favoring accurate NHEJ of short incompatible ends.

As shown in Fig. 4*A* (left), in the presence of a 5′-OH group the terminal transferase activity of human Pol$\mu$ extends the 1nt-protruding end with any of the four ddNTPs, ddCTP being the most efficient. However, the presence of a 5′P group (right) alters this specificity. Incorporation of ddTTP, ddCTP and ddGTP was significantly reduced (43%, 37%, and 46% decrease, respectively), but incorporation of ddATP was largely enhanced (163% increase). This phosphate-dependent behavior was specific of Pol$\mu$, being absent in TdT (Fig. S3). These results can be explained as indicated in Fig. 4*B*. If a 5′P is not present (scheme 1), Pol$\mu$ inserts untemplated nucleotides (N) by terminal transferase, assisted by its loop 1 (13). If a 5′P group is present (scheme 2), the preferred insertion of ddATP indicates that a single Pol$\mu$ molecule was able to bridge two incompatible DNA ends, mediated by loop 1, and copy the templating base (T) provided by a second end, thus achieving a precise NHEJ reaction, as it was shown to occur at different incompatible ends also when Pol$\mu$ was in the presence of NHEJ core factors (15). Bridging of the second end is mediated by interaction with its 5′P through the 8-kDa domain (Fig. 4*B*, schemes 2 and 3); the rationale for preferring the correct nucleotide (ddATP) is that the rate limiting step, that is, the movement of the protruding primer from an inactive to an active conformation, is facilitated by the approaching end, which provides a templating base "in trans." If this templating base fails to be located at a proper register, the ability of Pol$\mu$ to catalyze template-independent additions will allow joining (as represented in Fig. 4*B*, scheme 3), as Ligase IV can accept a mismatched 3′-end for ligation (20). These results indicate that a recessive 5′P is key for the precise joining of incompatible ends. Interestingly, and in agreement with recent findings supporting that 1 nucleotide is the optimal 3′-protrusion size for the NHEJ capacity of Pol$\mu$ on incompatible ends (15), the effect of the 5′P is not seen in longer protrusions (Fig. S4). When using longer 3′-protruding substrates, the enzyme only inserts untemplated nucleotides as is unable to accommodate the DNA ends.

**The Excessive Terminal Transferase of Mutant R387K Affects Fidelity during NHEJ of Incompatible Ends.** As expected from its enhanced terminal transferase activity and facilitated rate-limiting step, mutant R387K showed quantitative but also qualitative differences when compared to the wt enzyme on incompatible 3′-protruding ends. To show this, we used a different set of 3′-protruding substrates, either with compatible (Fig. 5*A*) or incompatible (Fig. 5 *B* and *C*) ends. On compatible ends, both wt Pol$\mu$ and R387K preferentially inserted the correct (templated) nucleotide (ddGTP), indicating that accurate NHEJ of minimally compatible ends has occurred (Fig. 5*A*). Their similar efficiency mimicks their behavior in gap-filling (Fig. 2*A*). As firstly shown by Davies et al. (15) the particular sequence (and size of the protrusion) forming the incompatible ends conditions the outcome of the NHEJ reaction. Thus, on incompatible ends that cannot be accurately joined (Fig. 5*B*), R387K was more efficient than wt Pol$\mu$ in NHEJ, according to the enhanced terminal transferase and reduced activation barrier of this mutant, but keeping the reaction very inaccurate (Fig. 5*B*). On the other hand, simply by changing the protruding nucleotide (to a dC), incompatible ends can now be accurately joined by the wild-type Pol$\mu$ (Fig. 5*C*), that is, the reaction is preferentially template-directed (ddG insertion). In this case, mutant R387K showed an error-prone behavior, probably due to its facilitated rate-limiting step, i.e., excessive terminal transferase activity, that favors the untemplated nucleotide addition of any dNTP. These results suggest that $Arg^{387}$ residue, which acts as a brake for TdT activity in Pol$\mu$, could avoid generation of excessive variability during NHEJ.

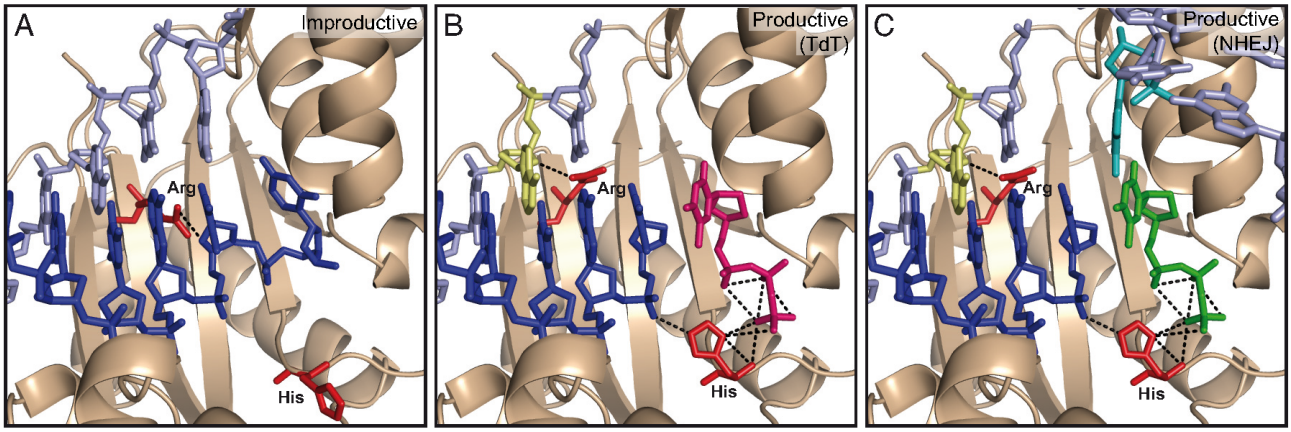



**Fig. 6.** Modeling the limiting step of Pol$\mu$'s terminal transferase. (*A*) Stable/non-productive step (binary complex): Pol$\mu$ (1IHM) was overimposed on the binary complex of TdT with ssDNA (1KDH). Only the partial structure of Pol$\mu$ (wheat color) is shown for clarity. A 1nt 3′-protruding substrate (derived from the gapped substrate present in the Pol$\mu$ crystal) was modeled (as in 1KDH) to reproduce the initial situation of a binary complex in which the primer (dark blue) occupies the NTP binding pocket. Residue His$^{329}$ (as His$^{342}$ in TdT) is in "standby," and Arg$^{387}$ (modeled from Lys$^{403}$ in TdT) strongly interacts with the primer strand. (*B*) Productive step for the terminal transferase (ternary complex in the absence of template): the primer (dark blue) has been relocated, and any incoming dNTP (magenta) sits in place with the assistance of His$^{329}$ (rotated 180°). Arg$^{387}$ stabilizes the new primer location by interacting with the template strand (nucleotide in yellow). (*C*) Productive step during NHEJ (ternary complex with a template provided in trans): a NHEJ reaction implying two 3′-protruding and incompatible ends (modeled from the gapped substrate present in the Pol$\mu$ crystal) is shown. Primer relocation and Arg$^{387}$ repositioning is facilitated by the template provided in trans (nucleotide in cyan) that allows selection of a complementary nucleotide (depicted in green).

## Discussion

In most DNA-dependent DNA polymerases, proper positioning of the 3′-terminus is indirectly dictated by the enzyme's avidity by the templating base, thus configuring a binary complex ready to select the incoming nucleotide (ternary complex). Eventually, when no template base is available (blunt or 3′-protruding ends), any further nucleotide addition is unfavoured, probably due to a deficient re-positioning of the 3′ terminus. Thus, only a poorly efficient + 1 addition reaction, i.e., a minimalist terminal transferase, can be displayed by some DNA polymerases lacking 3–5′ exonuclease (21–23). Conversely, TdT and Pol$\mu$ are adapted for template-independency in part due to a specific histidine residue (His$^{329}$ in Pol$\mu$ and His$^{342}$ in TdT) that actively cooperates in the relocation of the primer from a "non-productive" E:DNA complex to a "productive" E:DNA:dNTP complex in the absence of a templating base.

Combined structural and functional evidences for both Pol$\mu$ and TdT (17, 18), together with the functional data presented in this paper, indicate that there is another residue directly implicated in the terminal transferase activity of both enzymes. That residue (Arg$^{387}$ in Pol$\mu$ and Lys$^{403}$ in TdT) modulates the catalytic efficiency of the terminal transferase reaction, by regulating the rate-limiting step. In the case of Pol$\mu$, Arg$^{387}$ acts as a brake for the necessary movement of the primer, to limit nucleotide additions before end bridging. However, when a templating base is provided in trans during NHEJ, the rate-limiting step is relieved, and Arg$^{387}$ changes its blocking interaction with the primer for an interaction with the template strand (−3 position), to stabilize the productive complex, thus allowing efficient and accurate end-joining to occur (see Fig. 6). That interaction of Arg$^{387}$ cannot occur in mutant R387A, in agreement with its defective terminal transferase, and it would be lost at 3′-protruding ends longer than three nucleotides. Interestingly, mutant R387K produced a very specific blockage at position + 4 when continuous terminal transferase extension of a blunt end was tested (Fig. 2*B*). This suggests that the mutant lysine cannot properly establish the interactions occurring at this transition, when Arg$^{387}$ could stablish alternative interactions (possibly with loop1) on longer protrusions. Interestingly, the equivalent residue in human Pol$\lambda$ (Lys$^{472}$) is also involved in regulating the catalytic cycle by means of inhibitory interactions with the primer strand (T. A. Kunkel and K. Bebenek, personal communication). In the case of TdT, residue Lys$^{403}$ likely establishes a weaker interaction with the primer compared to its orthologue Arg$^{387}$ in Pol$\mu$. Thus, TdT has been optimized to efficiently overcome the rate-limiting step of the terminal transferase, to exclusively perform creative synthesis.

Which is the physiological role of the terminal transferase activity of Pol$\mu$? As shown here, and in other reported studies, NHEJ of short incompatible ends can be accurate in many cases, but imprecise in others depending on both the length and sequence of each protrusion (15). For the latter cases, when a templating base is not in a proper register, untemplated terminal transferase addition in a NHEJ context provides a valid, although mutagenic, solution that would be conceptually similar to translesion DNA synthesis. Besides, it cannot be ruled out that Pol$\mu$'s terminal transferase can extend a single short 3′-protrusion to facilitate end-joining of this fraction of non-complementary ends.

Based on our mutational data, we elaborated a model that explains why Pol$\mu$ prefers to have a more limited/mild terminal transferase activity, regulated by Arg$^{387}$. A templating base provided in trans by the approaching end that could be located in a proper register will stabilize the incoming (and complementary) nucleotide, thus facilitating primer translocation. As a result of this, NHEJ of many incompatible ends can be efficient and accurate. During NHEJ of this fraction of incompatible ends, as shown in this paper, an excessive terminal transferase would be disadvantageous in terms of genome stability. On the other hand, our findings also explain the necessity of a mild terminal transferase activity in Pol$\mu$, not only devoted to create connectivity in those other DNA ends that cannot be efficiently joined on a templating basis, but perhaps contributing to gain a certain degree of genome variability. Additionally, it can be inferred that TdT evolved to maximize the efficiency of the translocation mechanism in the absence of template, at the cost/benefit of introducing untemplated nucleotides, thus being devoted to generate variability at V(D)J recombination intermediates.

Is there any in vivo evidence of untemplated insertions made by Pol$\mu$? It has been shown that mice that are TdT−/− still

contain a 5% of their V(D)J junctions with template-independent additions, which suggested a possible role of Pol$\mu$ in these reactions (24). In agreement with that, the terminal transferase activity of Pol$\mu$ has been directly implicated at variability/repair processes occurring at embryo developmental stages in which TdT is still not expressed (25).

## Materials and Methods

**DNA and Proteins.** Synthetic DNA oligonucleotides were obtained from Isogen. PAGE-purified oligonucleotides were labeled at their 5′ ends with [$\gamma$-$^{32}$P]ATP. The oligonucleotides used to generate the DNA substrates were the following: for gapped substrates, P15 (5′-TCTGTGCAGGTTCTT-3′), T32 (5′-TGAAGTCCCTCTCGACGAAGAACCTGCACAGA-3′) and D16 (5′-GTCGAGAGGGACTTCA-3′); as ssDNA, PolydT (5′-TTTTTTTTTTTTTTTTTTTT-3′), to avoid the formation of secondary structures: for terminal transferase activity assays, oligonucleotides 5′-CGCAAGTCAGCGCTACGGG-3′ (Blunt), 5′-CGCAAGTCAGCGCTACGGGT-3′ (1T), 5′-CGCAAGTCAGCGCTACGGGTT-3′ (2T), 5′-CGCAAGTCAGCGCTACGGGTTT-3′ (3T), and 5′-CGCAAGTCAGCGCTACGGGTTTTTTTTT-3′ (9T) were used as primer strands, hybridized to oligonucleotide 3′-GCGTTCAGTCGCGATGCCC-5′ to obtain either blunt or 3′-protruding dsDNA substrates. For NHEJ assays, oligonucleotides 5′-CCCTCCCTCCCCA-3′ (CA), 5′-CCCTCCCTCCCGT-3′ (GT), 5′-CCCTCCCTCCCG-3′ (G) and 5′-CCCTCCCTCCCC-3′ (C) were used as primers, hybridized to oligonucleotide 5′-GGGAGGGAGGG-3′. Oligonucleotides D16, CA, GT, G, and C contain a phosphate at the 5′-end. Ultrapure dNTPs, ddNTPs, [$\alpha$-$^{32}$P] dNTPs (3,000 Ci/mmol) and [$\gamma$-$^{32}$P] ATP (3,000 Ci/mmol) were purchased from GE Healthcare. T4 polynucleotide kinase was obtained from New England Biolabs. Pfu DNA polymerase was purchased from Promega Corporation.

**Construction and Purification of Human Pol$\mu$ Mutant Proteins.** Site-directed mutagenesis (QuikChange Site Directed Mutagenesis kit, Stratagene) was performed on the overexpression plasmid for Pol$\mu$ (pRSETA-hPol$\mu$) (6), with primers: H329G (5′-AAGTTGCAGGGCGGAGACGTGGACTTC-3′ and 5′-GAAGTCCACGTCTCCGCCCTGCAACTT-3′), R387K (5′-GACGCTTTTGAGAAGAGTTTCTGCATT-3′ and 5′-AATGCAGAAACTCTTCTCAAAAGCGTC-3′) and R387A (5′-GACGCTTTTGAGGCGAGTTTCTGCATT-3′ and 5′-AATGCAGAAACTCGCCTCAAAAGCGTC-3′). DNA constructs were sequenced and transformed in *E. coli* BL21(DE3)pLysS. Wild-type and mutant Pol$\mu$ variants were overexpressed and purified as described (6).

**DNA Polymerization Assays.** To analyze DNA-dependent and independent polymerization, several DNA substrates, containing 5′P-labeled primers, were incubated with different proteins, at the concentration indicated in each case. The reaction mixture, in 20 $\mu$L, contained 50 mM Tris-HCl (pH 7.5), 1 mM DTT, 4% glycerol, and 0.1 mg/mL BSA, in the presence of the indicated amounts of the DNA polymerization substrates, and the indicated concentrations of dNTP and activating metal ions. After incubation, reactions were stopped by adding gel loading buffer [95% (vol/vol) formamide, 10 mM EDTA, 0.1% (wt/vol) xylene cyanol, and 0.1% (wt/vol) bromophenol blue] and analyzed by 8 M urea/20% PAGE and autoradiography. When indicated, we used ddNTPs instead of dNTPs to limit incorporation to a single nucleotide on the 3′-end of the labeled oligonucleotide. For a steady state analysis of the terminal trasferase activity, the reaction mixture, in 20 $\mu$L, contained 50 mM Tris-HCl (pH 7.5), 1 mM DTT, 4% glycerol, and 0.1 mg/mL BSA, in the presence of $MnCl_2$ (1 mM), polydT (1 $\mu$M), either Pol$\mu$ wt or R387K mutant (200 nM), and various concentrations of the indicated [$\alpha$-$^{32}$P] dNTP (10–1,000 nM). After incubation for 30 min at 30 °C, reactions were stopped with EDTA (0.35 M), filtered through Sephadex G-25 spun columns (Roche), and the Cerenkov radiation of the excluded volume, quantified. The plotted data were fitted by a non-linear regression curve to the Michaelis-Menten equation: $V = V_{max}*(dNTP)/[(K_m+(dNTP)]$. f$^{o}$ext, a constant that represents the ratio of the catalytic efficiencies, for each nucleotide, relative to Pol$\mu$ wt.

**Analysis of the Interaction of Pol$\mu$ with Single-Stranded DNA by Nitrocellulose Filter Binding.** The assay was carried out using 5 nM of a labeled homopolymer (PolydT) as DNA primer, containing a ddCMP in the 3′-end to avoid polymerization, in the presence of 1 mM $MnCl_2$, and the indicated concentration of either Pol$\mu$ wt, R387K, R387A, or H329G. When indicated, 100 $\mu$M dTTP was added. Reactions were incubated for 10 min at 22 °C in binding buffer (50 mM Tris·HCl, pH 8.0, 25 mM NaCl, and 1 mM DTT), and filtered on nitrocellulose filters as described (19). Radiolabeled DNA retained in each filter was quantitated by a PhosphorImager.

**Amino Acid Sequence Comparisons and 3D Modeling.** A multiple alignment of different DNA polymerases of the Pol X family was done using the program MULTALIN (http://prodes.toulouse.inra.fr/multalin/). PDB coordinates for two different TdT structures and for one Pol$\mu$ structure, obtained from the Protein Data Bank (http://www.rcsb.org/pdb/), were used: 1KDH (binary complex of murine TdT with ssDNA); 1KEJ (murine TdT complexed with ddATP); 1IHM (ternary complex of murine Pol$\mu$ with gapped-DNA and incoming ddNTP). The different conformations of the studied residues of Pol$\mu$ and TdT were analyzed by using the Swiss PDB Viewer (http://www.expasy.ch/spdbv/) and MacPymol (http://delsci.com/macpymol/) programs.

**ACKNOWLEDGMENTS.** We thank M. de Vega, T. A. Kunkel, and K. Bebenek for helpful discussions and critical reading of the manuscript. This work was supported by Ministerio de Ciencia y Tecnología Grants BFU2006–14390/BMC, CONSOLIDER CSD2007–00015 and Comunidad Autónoma de Madrid Grants P2006/BIO-0306 to L.B., and by an institutional grant to Centro de Biología Molecular "Severo Ochoa" from Fundación Ramón Areces. M.J.M. was recipient of a fellowship from Comunidad Autónoma de Madrid. J.F.L de S. is an Investigator of the Ramon y Cajal Program.

1. Friedberg EC, et al. (2006) in *DNA Repair and Mutagenesis* (ASM Press, Washington D.C.), 2nd Ed.
2. Bebenek K, Kunkel TA (2004) Functions of DNA polymerases. *Adv Protein Chem* 69:137–165.
3. Moon AF, et al. (2007) The X family portrait: Structural insights into biological functions of X family polymerases. *DNA Repair (Amst)* 12:1709–1725.
4. Bork P, et al. (1997) A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J* 11:68–76.
5. Callebaut I, Mornon JP (1997) From BRCA1 to RAP1: A widespread BRCT module closely associated with DNA repair. *FEBS Lett* 400:25–30.
6. Dominguez O, et al. (2000) DNA polymerase mu (Pol mu), homologous to TdT, could act as a DNA mutator in eukaryotic cells. *EMBO J* 19:1731–1742.
7. García-Díaz M, et al. (2000) DNA polymerase lambda (Pol lambda), a novel eukaryotic DNA polymerase with a potential role in meiosis. *J Mol Biol* 301:851–867.
8. Ruiz JF, et al. (2001) DNA polymerase mu, a candidate hypermutase? *Philos Trans R Soc London B Biol Sci* 35:699–709.
9. Mahajan KN, et al. (1999) Association of terminal deoxynucleotidyl transferase with Ku. *Proc Natl Acad Sci USA* 96:13926–13931.
10. Mahajan KN, Nick McElhinny SA, Mitchell BS, Ramsden DA (2002) Association of DNA polymerase mu (pol mu) with Ku and ligase IV: Role for Pol mu in end-joining double strand break repair. *Mol Cell Biol* 22:5194–5202.
11. Ma Y, et al. (2004) A biochemically defined system for mammalian nonhomologous DNA end joining. *Mol Cell* 16:701–713.
12. Nick McElhinny SA, et al. (2005) A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Mol Cell* 19:357–366.
13. Juarez R, Ruiz JF, Nick McElhinny SA, Ramsden D, Blanco L (2006) A specific loop in human DNA polymerase $\mu$ allows switching between creative and DNA-instructed synthesis. *Nucleic Acids Res* 34:4572–4582.
14. Ruiz JF, et al. (2004) Overexpression of human DNA polymerase $\mu$ (Pol $\mu$) in a Burkitt's lymphoma cell line affects the somatic hypermutation rate. *Nucleic Acids Res* 32:5861–5873.
15. Davies BJ, Havener JM, Ramsden DA (2008) End-bridging is required for Pol $\mu$ to efficiently promote repair of noncomplementary ends by nonhomologous end joining. *Nucleic Acids Res* 36:3085–3094.
16. Tonegawa S (1983) Somatic generation of antibody diversity. *Nature* 302:575–581.
17. Delarue M, et al. (2002) Crystal structures of a template-independent DNA polymerase: Murine terminal deoxynucleotidyl transferase. *EMBO J* 21:427–439.
18. Moon AF, et al. (2007) Structural insight into the substrate specificity of DNA Polymerase mu. *Nat Struct Mol Biol* 14:45–53.
19. Wong I, Lohman TM (1993) A double-filter method for nitrocellulose-filter binding: Application to protein-nucleic acid interactions. *Proc Natl Acad Sci USA*, 90:5428–5432.
20. Gu J, et al. (2007) XRCC4:DNA ligase IV can ligate incompatible ends and can ligate across gaps. *EMBO J* 26:1010–1023.
21. Clark JM, Joyce CM, Beardsley GP (1987) Novel blunt-end addition reactions catalyzed by DNA polymerase I of *Escherichia coli*. J Mol Biol 198:123–127.
22. Clark JM (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res* 16:9677–9686.
23. Golinelli MP, Hughes SH (2002) Nontemplated nucelotide addition by HIV-1 reverse transcriptase. *Biochemistry* 41:5894–5906.
24. Bertocci B, De Smet A, Weill JC, Reynaud CA (2006) Nonoverlapping functions of DNA polymerases $\mu$, $\lambda$, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination in vivo. *Immunity* 25:31–41.
25. Gozalbo-López B, et al. (2009) A role for DNA polymerase mu in the emerging DJH rearrangements of the postgastrulation mouse embryo. *Mol Cell Biol* 29:1266–1275.

Molecular Cell
Article

# Structure of a Preternary Complex Involving a Prokaryotic NHEJ DNA Polymerase

Nigel C. Brissett,[1,4] Maria J. Martin,[2,4] Robert S. Pitcher,[1] Julie Bianchi,[1] Raquel Juarez,[2] Andrew J. Green,[1] Gavin C. Fox,[3] Luis Blanco,[2,*] and Aidan J. Doherty[1,*]
[1]Genome Damage and Stability Centre, University of Sussex, Brighton BN1 9RQ, UK
[2]Centro de Biología Molecular Severo Ochoa, CSIC-UAM, 28049 Madrid, Spain
[3]Synchrotron Soleil, Ormes des Merisiers, 91190 Saint Aubin, France
[4]These authors contributed equally to this work
*Correspondence: lblanco@cbm.uam.es (L.B.), ajd21@sussex.ac.uk (A.J.D.)
DOI 10.1016/j.molcel.2010.12.026

## SUMMARY

**In many prokaryotes, a specific DNA primase/polymerase (PolDom) is required for nonhomologous end joining (NHEJ) repair of DNA double-strand breaks (DSBs). Here, we report the crystal structure of a catalytically active conformation of *Mycobacterium tuberculosis* PolDom, consisting of a polymerase bound to a DNA end with a 3′ overhang, two metal ions, and an incoming nucleotide but, significantly, lacking a primer strand. This structure represents a polymerase:DNA complex in a preternary intermediate state. This polymerase complex occurs in solution, stabilizing the enzyme on DNA ends and promoting nucleotide extension of short incoming termini. We also demonstrate that the invariant $Arg^{220}$, contained in a conserved loop (loop 2), plays an essential role in catalysis by regulating binding of a second metal ion in the active site. We propose that this NHEJ intermediate facilitates extension reactions involving critically short or noncomplementary DNA ends, thus promoting break repair and minimizing sequence loss during DSB repair.**

## INTRODUCTION

Two major cellular pathways have evolved to repair DSBs: nonhomologous end joining (NHEJ) and homologous recombination (HR) (Daley et al., 2005; Mahaney et al., 2009). In HR, an intact duplex, homologous to the break site, acts as an alternative copy that templates for the resynthesis of the broken strands (Helleday et al., 2007). In contrast, NHEJ does not require extensive homology but instead directly joins DNA ends together in a template-independent manner. Consequently, NHEJ can be utilized by noncycling cells, when the absence of a homologous donor precludes double-strand break (DSB) repair by HR, and, for this reason, it has become the major pathway for repairing DSBs in mammalian cells (Helleday et al., 2007).

The higher eukaryotic NHEJ complex consists of Ku, DNA-PKcs, and the ligase IV:XRCC4:XLF (LXX) complex (Daley et al., 2005; Mahaney et al., 2009). Upon formation of a DNA break in cells, Ku and DNA-PKcs bind with high affinity to the ends of DSBs, maintaining the termini in close proximity. Subsequently, Ku recruits the LXX complex, thereby directing ligase IV activity. A host of other proteins have also been implicated in eukaryotic NHEJ, including DNA polymerases λ and μ (PolX family), the nuclease Artemis, and PNKP (polynucleotide kinase and phosphatase). These additional factors are involved in the replacement, removal, or modification of damaged or incompatible DNA termini to produce ligatable DNA ends.

More recently, a functionally homologous NHEJ repair apparatus has been characterized in prokaryotes (Weller et al., 2002; Della et al., 2004; Gong et al., 2005; Pitcher et al., 2007a). The *Mycobacterium tuberculosis* (*Mt*) NHEJ complex, composed of Ku and a multifunctional DNA ligase D (LigD), can recognize, process, and reseal DSBs (Della et al., 2004; Pitcher et al., 2005; Pitcher et al., 2007b). This primordial NHEJ complex is present in many bacterial species, suggesting that this repair pathway is physiologically important for genome stability in prokaryotes (Bowater and Doherty, 2006). NHEJ-deficient stationary phase cells (*Mycobacteria*) and spores (*Bacillus*) are sensitive to agents that produce DSBs (Weller et al., 2002; Moeller et al., 2007; Pitcher et al., 2007c), indicating that NHEJ is also predominantly utilized by nonreplicating prokaryotic cells.

The mycobacterial LigD, in addition to its ligase domain, also contains polymerase (PolDom) and nuclease (NucDom) domains, which can process DSBs prior to ligation (Della et al., 2004). The prokaryotic NHEJ polymerases (PolDom) are members of the Archaeo-Eukaryotic Primase (AEP) superfamily (Aravind and Koonin, 2001; Weller and Doherty, 2001; Iyer et al., 2005). Historically, primases have been regarded as specialized polymerases with roles restricted to the synthesis of RNA primers at replication forks (Frick and Richardson, 2001). However, this limited role for AEPs in DNA metabolism has been challenged by the discovery of a wide variety of structurally homologous bacterial AEPs (Pitcher et al., 2007a) that possess multiple polymerase activities, including gap-filling activities (Della et al., 2004; Zhu and Shuman, 2005; Zhu et al., 2006; Pitcher et al., 2007b). In addition to these activities, a reduced template-dependence and flexible active site enables these enzymes to generate template distortions, primer realignment, and lesion bypass, which greatly enhances their roles in DSB repair (Pitcher et al., 2007b).

Molecular Cell
Structure of a NHEJ Polymerase Preternary Complex

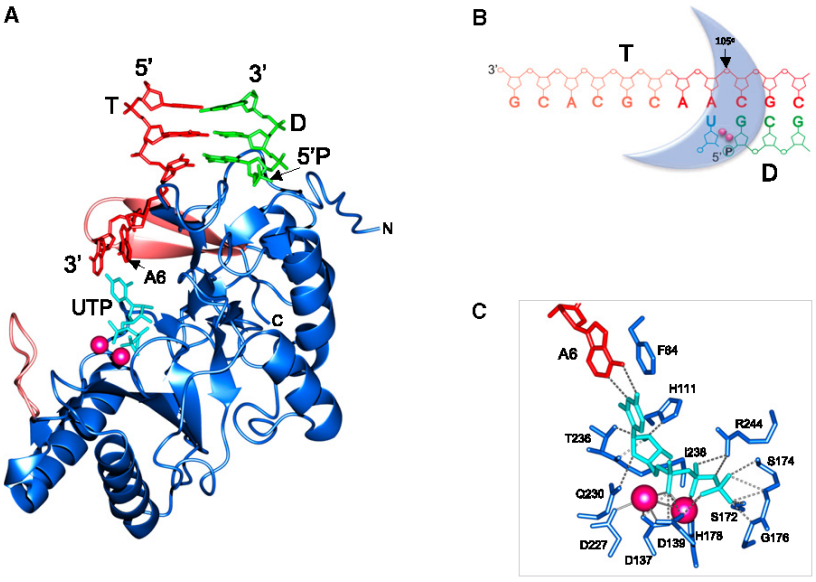



**Figure 1. Crystal Structure of the Preternary Complex of *Mt*-PolDom:DNA:UTP**

(A) Ribbon representation of the preternary complex depicting the interaction of PolDom (blue) with the dsDNA and UTP. Loops 1 and 2 are highlighted in salmon pink showing their spatial relationship to the bound DNA and UTP molecules. DNA is colored green for downstream (D) strand and red for templating (T) strand. The bound UTP molecule is colored cyan and the catalytic manganese ions are pink.

(B) Schematic representation of *Mt*-PolDom binding to DNA and UTP. The unstructured 3′-overhanging DNA is shown as a faded color as it is unstructured. DNA chains follow the same color scheme as in (A). PolDom is represented by a blue crescent. The arrow indicates a kink angle (105°) between A6 and C5 of the template strand.

(C) A close-up view of the active site of *Mt*-PolDom highlighting the important stacking of A6 against Phe$^{64}$ and the specific interactions of PolDom with UTP and UTP with DNA (A6). The DNA and protein monomer are colored as in (A). Hydrogen bonds are shown in gray.

See also Figure S1.

Uniquely among the domains that constitute *Mt*-LigD, PolDom physically interacts with Ku (Pitcher et al., 2005), suggesting that LigD recruitment to DSBs is accomplished primarily via this interaction. Ku also appears to enhance polymerase extension activity (Pitcher et al., 2005). PolDom can also facilitate the association (synapsis) of noncomplementary DNA ends, thus promoting end resection prior to gap filling and ligation (Brissett et al., 2007). Here, we provide structural and functional evidence for the existence of a catalytically competent NHEJ polymerase:DNA preternary complex involved in the repair of DNA DSBs. The crystal structure of the *Mt*-PolDom in complex with double-stranded DNA (dsDNA), containing a 3′ overhang, reveals how these polymerases achieve templated binding of an incoming nucleotide (UTP) in the active site, in the presence of divalent metal ions, and the localized structural transitions associated with this step. Biochemical and biophysical data are presented that support the existence of this catalytic intermediate in solution. The critical role of an invariant arginine residue (Arg$^{220}$), located in a conserved loop (loop 2), in regulating metal binding and catalysis is also reported. Formation of a preternary complex enhances both the binding of the polymerases to DNA ends and subsequent nucleotide extension of short incoming termini. We propose that the preternary polymerase complex represents a functional NHEJ intermediate that facilitates joining of short noncomplementary DNA ends and, consequently, minimizes sequence loss during DSB repair.

## RESULTS

### Structure of a PolDom:DNA:UTP Preternary Complex

Here, we present the crystal structure of *Mt*-PolDom in complex with dsDNA (with an 8 base 3′ overhang and a recessed 5′ P) and bound incoming nucleotide (UTP) and manganese ions located in the active site. This *Mt*-PolDom complex (PolDom:UTP:$Mn^{2+}$:DNA) represents an incomplete ternary complex, lacking a primer strand, and is therefore ascribed as a preternary complex (Figure 1). The overall complex in the asymmetric unit consists of two distinct preternary complexes and specific details of the DNA and nucleotide interactions with both monomers are described in the Supplemental Results (Figures S1 and S2 available online).

The PolDom structure comprises amino acid residues 6–291, with no electron density observed for nine residues at the C terminus (Table 1). PolDom interacts with the recessed 5′ phosphate of the downstream strand (D) (Figure 1 and Figure S1) and orients the 3′-overhanging end of the templating strand (T), which is splayed out by ~105° (Figure 1B). Only two bases from the 3′-overhanging strand have observable density in the refined structure, suggesting that the remaining six bases are highly mobile. A single UTP is bound in the polymerase active site, which forms a Watson-Crick base pair with a templating adenine base (A6) with two metals ions also bound in close proximity (Figure 1C). The 3′ OH group of the ribose moiety of UTP is hydrogen bonded to the amino nitrogen of Ile$^{238}$. The 2′ OH group is hydrogen bonded to the side-chain hydroxyl and backbone carboxyl oxygen of Thr$^{236}$, with a further hydrogen bond formed between the δN of His$^{111}$ and the 2′ OH. Gln$^{230}$ interacts with the oxygen (O4′) on the ribose moiety (Figure 1C). To our knowledge, this is the only reported structure of a polymerase-DNA complex containing a templated incoming base in the absence of a primer strand, a scenario that is most relevant for NHEJ polymerases. The biochemical and structural significance of this complex is described below.

### Structural Evidence Supporting a Precatalytic Ternary Complex

A comparison of the active sites of the binary complex (PolDom bound to GTP and metal) (Pitcher et al., 2007b) with the preternary complex revealed significant differences in the positioning of the nucleotides in the active sites. Notably, the position

Molecular Cell
Structure of a NHEJ Polymerase Preternary Complex

**Table 1. Data Collection and Refinement Statistics**

| | |
|---|---|
| Data Collection | |
| Source | ESRF BM16-1 |
| Space group | $P4_1$ |
| Unit cell dimensions | |
| *a*,*b*,*c* (Å) | 145.71/145.71/44.96 |
| α,β,γ | 90.000/90.000/90.000 |
| Wavelength (Å) | 0.9787 |
| Resolution (Å) | 50.00–3.10 |
| Total Number of observations | 97,102 |
| Number of unique reflections | 17,593 |
| Overall I/(σI)[a] | 16.6 (2.7) |
| Overall completeness (%)[a] | 99.4 (97.5) |
| $R_{sym}$ (%)[a,b] | 16.1 (67.9) |
| Redundancy[a] | 5.5 (4.4) |
| Refinement | |
| Resolution (Å) | 48.56–3.10 |
| Number of of reflections | 16,680 |
| $R_{factor}$ / $R_{free}$[c,d] | 0.1907/0.2476 |
| Contents of asymmetric unit | 6 molecules (2 protein, 2 DNA, 2 nucleotide) |
| Number of atoms | |
| Protein | 4377 |
| DNA/Nucleotide | 223/58 |
| Water molecules/metal ions | 47/4 |
| Mean B value (Å$^2$) | 49.09 |
| Rmsds | |
| Bonds (Å) | 0.01 |
| Angles (°) | 1.3 |
| Ramachandran statistics | |
| Favored regions (%) | 94.2 |
| Allowed regions (%) | 5.1 |
| Disallowed regions (%) | 0.7 |
| PDB accession code | 3PKY |

[a] Values for highest resolution shell (3.21–3.10 Å) is in parentheses.
[b] $R_{sym} = \Sigma|I - \langle I \rangle| / \Sigma \langle I \rangle$, where $I$ is the observed intensity.
[c] $R_{factor} = \Sigma||F_o| - |F_c|| / \Sigma|F_o|$, where $F_o$ and $F_c$ are the observed and calculated structure factor, respectively.
[d] $R_{free}$ is equal to R factor for a randomly selected 5% subset of reflections not used in the refinement.

of the triphosphate tail appears to dictate if a second metal ion, absent in the binary complex, can bind in the active site (Figure 2A). The orientation of this tail is strongly influenced by template:nucleotide pairing. In the preternary complex, the orientation of the templating adenine base (A6), forming a Watson-Crick base pair with the uridine base, is determined by a close stacking interaction with Phe$^{64}$ (Figure 1C). Significantly, the incoming nucleotide base in the GTP binary complex structure (with no templating base available) directly stacks against this phenylalanine (Figure 2A), changing its overall location relative to that observed in the preternary complex. Thus, Phe$^{64}$ (invariant in NHEJ AEPs) plays a pivotal role in the sequential recognition and orientation of both the templating and incoming bases (Pitcher et al., 2007b) and, therefore, also has an important role in determining the enzyme's affinity for a second active site metal ion.

Further support for the catalytic competency of the preternary polymerase complex is provided by comparisons of the conformations of UTP (preternary structure) with the Polλ ternary complex (dUpnpp, PDB ID: 2PFO; Garcia-Diaz et al., 2007), which revealed a pairwise atom alignment of 0·25Å, suggesting that UTP (in the preternary complex) is in a catalytically competent conformation (Figure 2B). The lack of an incoming priming strand means that phosphotransfer is unable to occur and therefore, the current PolDom structure must be defined as a preternary catalytic complex. However, the conformation of residues and cofactors (UTP and metals) in the active site suggests that the PolDom complex is optimally poised for extension upon engagement with an incoming primer strand.

## Movement of Loop 2 in the Preternary Complex

We next examined whether any major structural rearrangements had occurred in the PolDom:DNA complex upon templated binding of UTP and metal ions in the active site. Notably, the polymerase in the preternary complex adopts a similar overall conformation to that observed in the apo, nucleotide, and synaptic PolDom structures (Brissett et al., 2007; Pitcher et al., 2007b), with only minor local conformational changes (see the Supplemental Results for details). The root mean square deviation (rmsd) values are all below 1.0 Å, suggesting no major global conformational differences between the various polymerase structures. However, despite these overall similarities, a closer examination of the structural elements within the preternary PolDom complex revealed that a specific conserved loop (loop 2; residues 213–223) showed large rmsd values compared to the equivalent region in other PolDom structures (*apo*, GTP, dGTP, and synaptic complex). Loop 2 exists as a $3_{10}$ helix in all of the previously determined structures. However, this helix unravels and adopts a random coil conformation in the preternary complex (Figure 2C). This significant conformational change in loop 2 results in Cα position shifts of ~6 Å, inducing repositioning of specific side chains on the loop. A comparison of the amino acid sequences of PolDom's revealed that loop 2 is conserved across the entire family of NHEJ primases (Figure 2D), but not in the replicative orthologs, suggesting that this structural element could have a role in NHEJ. Closer inspection of the sequence of loop 2 identified an invariant arginine residue (Arg$^{220}$). Notably, the side chain of Arg$^{220}$ is orientated away from the active site, upon transition from the binary to the preternary complex (Figure 2C), suggesting that Arg$^{220}$ may play a specific role in catalysis. The functional significance of the movement of loop 2 is described below.

## Repositioning of Loop 2 Regulates Metal Binding in the Polymerase Active Site

Does this major reorientation of loop 2 away from the active site, upon formation of the preternary complex, play any functional role in the catalytic cycle of these NHEJ polymerases? Figures 3A and 3B show views of the active site of PolDom-DNA binary

Molecular Cell
Structure of a NHEJ Polymerase Preternary Complex

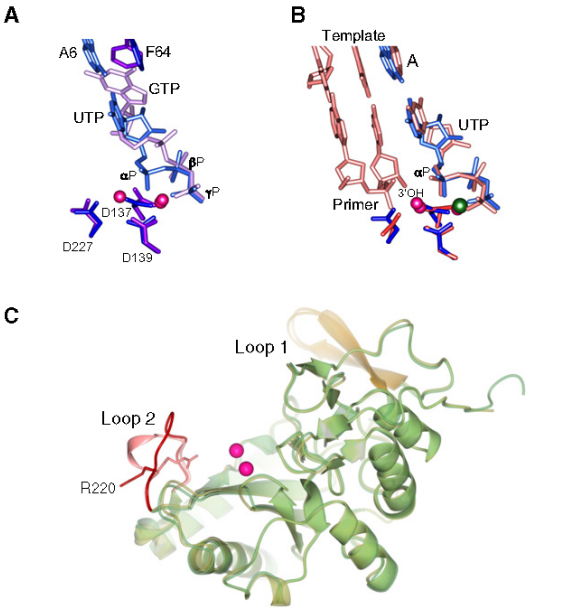



**Figure 2. Structural Evidence Supporting a Precatalytic Ternary Complex**

(A) Schematic view of the changes affecting nucleotide positioning, comparing the binary (*Mt*-PolDom:GTP; PDB ID:2IRX) complex (lilac) versus the preternary complex (blue). Metal ions (A and B) are indicated as pink spheres. Rotation of $Phe^{64}$ to establish alternative interactions either with the incoming nucleotide (GTP) or with the templating base (A6) is highlighted.

(B) UTP is in a catalytically competent conformation in the preternary complex. Superposition of the catalytic site of Polλ (PDB ID: 2PFO) (Garcia-Diaz et al., 2007) onto the catalytic site of the PolDom preternary complex. Catalytic aspartates from *Mt*-PolDom are colored dark blue with the UTP and templating base from the same structure colored light blue. Catalytic aspartates from the Polλ structure are colored red and the DNA, 2′-deoxyuridine 5′-alpha, beta-imido-triphosphate (dUPnPP) colored salmon pink. $Mn^{2+}$ ions are colored pink, and $Mg^{2+}$ ions are colored green.

(C) Ribbon representation highlighting the conformational changes associated with loop 2, largely affecting the orientation of $Lys^{217}$ and $Arg^{220}$. *Mt*-PolDom structures corresponding to the binary (PDB ID: 2IRX) and preternary complexes were superposed. The two metal ions present in the preternary complex are indicated in deep pink. The region corresponding to loop 2, and residue $Arg^{220}$, are indicated either in red (binary) or in pink (preternary). Loop 1 is also depicted (yellow) to show the relative orientation of these structural elements.

(D) Amino acid sequence alignment of the PolDom region encompassing loop 2 in the two main groups of bacterial LigD enzymes (see Figure S3 for enzyme nomenclature, amino acid positions and a more extended alignment). Invariant or highly conserved amino acids are indicated in red and orange, respectively. Dots above the *Mt*-PolDom sequence indicate the catalytically relevant $Arg^{220}$ (studied in this paper) and the metal ligand carboxylate residue $Asp^{227}$. Below the alignment is the consensus sequence, represented as a sequence logo (Crooks et al., 2004), showing that $Arg^{220}$ is the most conserved residue in loop 2.

See also Figures S1 and S3.

complex (left panels) and the preternary PolDom complex (right panels) containing an incoming UTP base-paired to a templating adenine base (A6) with two metal ions also bound. A number of critical observations suggest that the preternary complex is catalytically competent and that loop 2 plays a critical role in activating the catalytic center. The first notable observation is that a salt bridge formed between $Asp^{139}$ and $Arg^{220}$ (loop 2), observed in the DNA binary complex (Figure 3A, left), is disrupted in the preternary complex (Figure 3A, right). This allows a rotation of the arginine side chain (~180°) away from the active site, enabling $Asp^{139}$ to adopt a conformation where it can favorably act as a ligand for the second catalytic metal ion (A), not present in the catalytically inactive structures (Pitcher et al., 2007b). Accordingly, mutation of this arginine to alanine (R220A) significantly reduced PolDom's ability to incorporate nucleotides (Figure 3C) but, significantly, did not alter its affinity for gapped-DNA substrates (data not shown). It appears that binding of this second metal ion triggers this conformational change and this significant rearrangement at the active site implies that $Arg^{220}$ may play a direct role in catalysis.

Based on both these structural and biochemical observations, we propose that Watson-Crick base pairing of an incoming nucleotide (forming a chelating site for metal B), complementary to the templating base nearest to the recessive 5′ P, acts as the primary conformational switch that induces the active site to become catalytically competent. Thus, after selection of the correct nucleotide, a second metal (A) is recruited to the active site, perhaps initially in a noncompetent location requiring the interaction between $Arg^{220}$ and $Asp^{139}$ to provide the necessary space/charge to bind this second metal ion. Subsequently, a metal-driven conformational change of loop 2 leads to the breakage of the $Arg^{220}$-$Asp^{139}$ salt bridge, enabling $Asp^{139}$ to act as a chelating metal-binding ligand to promote repositioning of metal A into a catalytically competent location within the active site. Therefore, the capacity to form a salt bridge between $Arg^{220}$ and $Asp^{139}$ (abolished in the R220A mutant), which keeps the active site in a "standby" mode, is an important requisite step preceding catalysis. This preactivated complex, loaded with a complementary nucleotide, can occur at a single DNA end before a second 3′-protruding end provides an attacking primer for nucleotide insertion and NHEJ.

## Formation of a Stable Preternary Complex in Solution

To support the structural evidence for a preternary polymerase complex, we measured the formation and stability of PolDom preternary complexes in solution, adopting established fluorescent methodologies used to analyze polymerase ternary complexes (Kumar et al., 2008). Formation of a preternary complex was assessed by adding UTP to a preformed fluorescent complex of *Mt*-PolDom:DNA:$Mn^{2+}$, in which the 3′-protruding DNA end contains a 2-aminopurine [2AP] templating base. Titration of the UTP concentration enabled the determination of a binding constant ($K_d$) for nucleotide in the preternary complex (*Mt*-PolDom:DNA:UTP:$Mn^{2+}$). We observed

Molecular Cell
Structure of a NHEJ Polymerase Preternary Complex

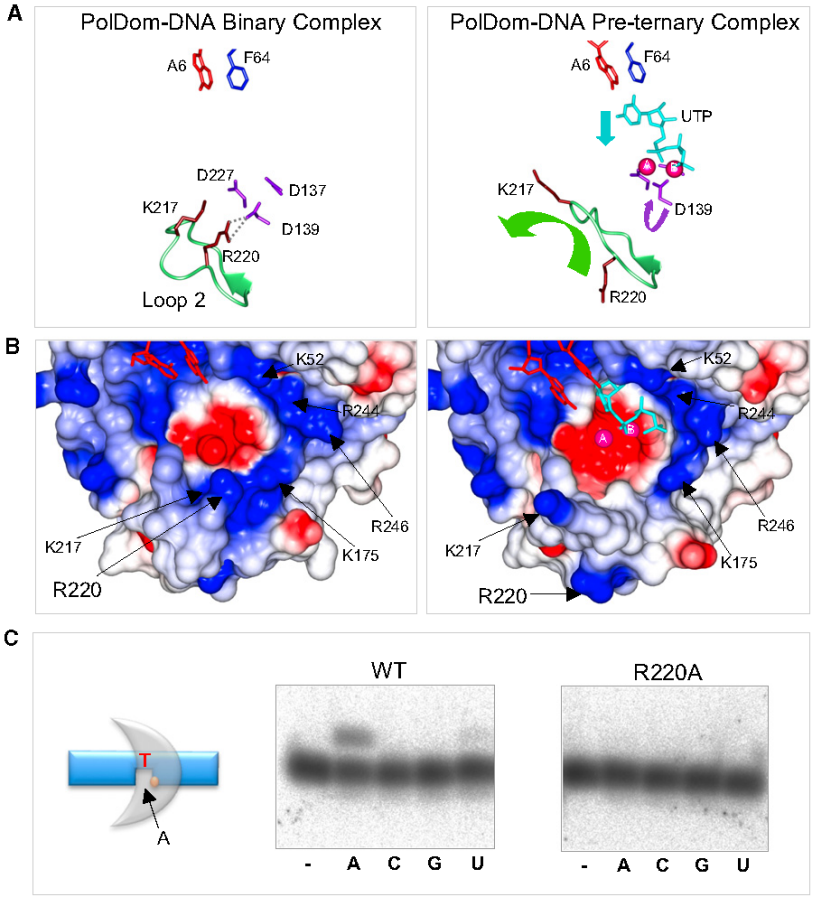



**Figure 3. Role of Loop 2 in Regulating the Active Site of PolDom**

(A) Schematic view of the conformational changes affecting loop 2 (green) and some selected residues acting as metal or nucleotide ligands, comparing the *Mt*-PolDom:DNA binary complex (PDB ID: 2R9L; left) versus the preternary complex (right). Metal ions (A and B) are indicated as pink spheres. Incoming nucleotide (UTP) is colored cyan. The templating nucleotide ($A_6$) present in the preternary complex, is indicated in red.

(B) Changes in electrostatic surface potential in the active site pocket, comparing the *Mt*-PolDom:DNA binary complex (PDB ID: 2R9L; left) versus the preternary complex (right). Loop 2 residues ($Lys^{217}$, $Arg^{220}$) are highlighted, as well as the residues involved in the formation of the positively charged ridge (in blue) above the active site ($Lys^{52}$, $Lys^{175}$, $Arg^{244}$, $Arg^{246}$).

(C) Gap-filling activity of *Mt*-PolDom loop 2 mutant, R220A. DNA synthesis reactions on a 5′ P-containing (yellow) 1 nt gapped substrate (5 nM) were carried out as described previously (Pitcher et al., 2007a), in the presence of either wild-type or mutant (R220A) *Mt*-PolDom (400 nM), 1 mM $Mn_2Cl$, and 100 μM of either the correct (ATP) or incorrect NTP. After incubation for 30 min at 30°C, extension of the 5′-labeled primer strand was analyzed by 8 M urea-PAGE (20% polyacrylamide gel) and autoradiography. See also Figure S3.

that a stable preternary PolDom complex was formed in solution, as pairing of UTP to the templating base quenched the intrinsic fluorescent signal of 2AP. Figure 4A shows the normalized fluorescence of the preformed *Mt*-PolDom:DNA:$Mn^{2+}$ complex with varying concentrations of UTP and the ensuing quenching that accompanies base pairing of 2-aminopurine with UTP. From these data, we calculated a $K_d$ of 0.646 μM for the binding of UTP to 2-aminopurine, determined by fitting relative quenching values to a ligand binding equation with a nonspecific binding factor. The submicromolar range $K_d$ for UTP binding closely correlates with the binding constants determined for nucleotide binding for other DNA polymerase ternary complexes (Kumar et al., 2008).

Next, we evaluated the formation and stability of both binary and preternary DNA complexes in solution, by electro-mobility shift assays (EMSAs) with a T/D DNA molecule with a short 3′ protrusion (GTT-3′), in the presence/absence of activating divalent metal ions and incoming nucleotides. As shown in Figure 4B, *Mt*-PolDom produced a stable binary complex with DNA in the absence of metal ions, observed as a sharp retarded band. Addition of either 10 μM $Mn^{+2}$ or 100 μM $Mg^{+2}$ did not significantly affect the formation and stability of such a binary complex. However, addition of both activating metal ions and CTP (the required nucleotide to form a correct preternary complex), significantly enhanced the stability of the complex formed. Conversely, addition of noncomplementary nucleotides had either an inhibitory effect (GTP; see Figure 4B) or a negligible effect (ATP; data not shown) on the stability of the polymerase preternary complex.

## Functional Significance of the Preternary Polymerase Complex

Although *Mt*-PolDom can form a stable preternary complex in solution, is this complex catalytically active? Moreover, is such a preformed complex beneficial for catalysis? To address these questions, we tested whether formation of a preternary complex, prior to the encounter with a primer, resulted in altered primer extension activity. For this, we used a trinucleotide to minimize terminal transferase-like nucleotide additions driven by *Mt*-PolDom. A schematic of this experiment is shown in Figure 5A. Figure 5B shows that the amount of 3G primer extended (dG-dG-dG-CMP) is significantly higher (~30% more efficient) when PolDom was preincubated with the templating molecule (…GCC-3′), correct nucleotide (CTP), and metal ions ($MnCl_2$) in the absence of primer, than when all of the reactants were added simultaneously. This beneficial effect of the preincubation did not occur when other combinations of cofactors were preincubated (Figure 5B). Therefore, the only situation in which a catalytic advantage was observed (see the reaction kinetics in Figure 5C) was after addition of the four components needed

Molecular Cell
Structure of a NHEJ Polymerase Preternary Complex

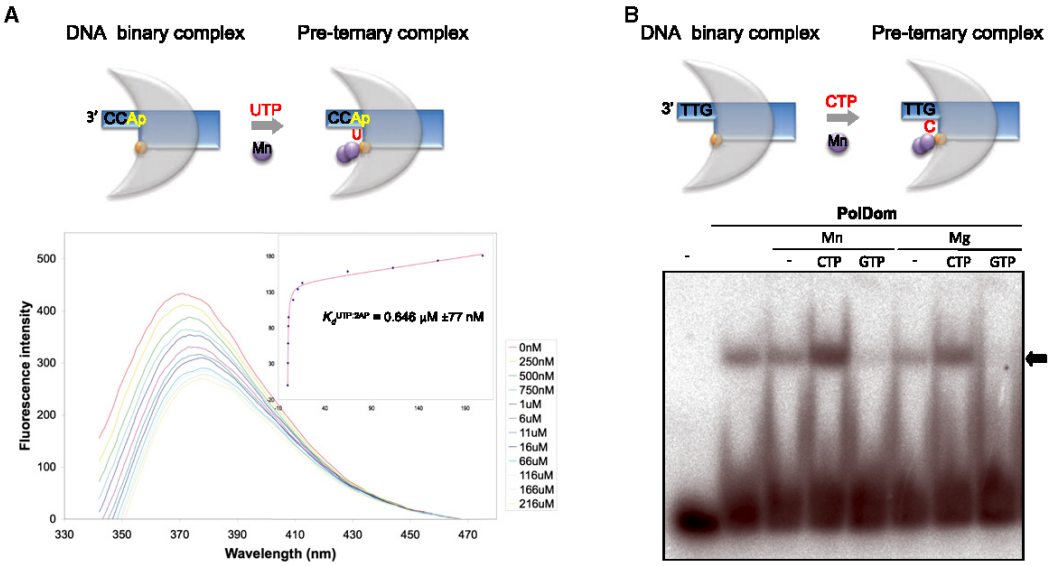



**Figure 4. Formation of a Stable Preternary Polymerase Complex in Solution**

(A) Titration of *Mt*-PolDom:2AP-labeled DNA complex with free UTP in the presence of $MnCl_2$. The main graph shows the normalized fluorescence of *Mt*-PolDom:2AP-labeled DNA complex (200 nM *Mt*-PolDom, 200 nM ddC terminated 2-aminopurine labeled DNA, 50 mM Tris [pH 7.5], 150 mM NaCl, 10 μM $MnCl_2$, and 4% glycerol) titrated with varying concentrations of UTP. Inset is a graph showing the relative fluorescence quenching as a function of UTP concentration (blue diamonds) overlaid with data fit to the nonlinear regression ligand binding equation (pink line).

(B) A stable preternary complex formed in solution. As detected by electrophoretic mobility shift assays (EMSA), *Mt*-PolDom binds 3′-protruding (…GTT-3′) T/D ends with a 5′ P to form a PolDom:DNA binary complex (arrow). Addition of activating metal ions and a complementary nucleotide (CTP) allows formation of a more stable preternary complex. Conversely, a stable preternary complex is not formed in the presence of a noncomplementary (GTP) nucleotide. EMSA assays were performed with 600 nM *Mt*-PolDom and 5 nM DNA, as described (Pitcher et al., 2007b). When indicated, cofactors (100 μM $MgCl_2$, 10 μM $MnCl_2$,) and nucleotides (either CTP or GTP 100 μM) were added.

to form a preternary complex. Similar results were also obtained with primers of different lengths and matching different DNA ends (data not shown).

### Preternary Complex Formation Enhances Extension of Short Incoming Termini

The results described above support the catalytic significance of a preternary complex in solution and suggest that such complexes enable these enzymes to extend off otherwise unstable termini (e.g., very short or blunt termini) or to deploy additional extension activities (e.g., terminal transferase). In the preternary complex, the position of a number of conserved positively charged residues (e.g., $Lys^{175}$ and $Arg^{246}$), which line the active site pocket, has altered significantly with respect to the PolDom:DNA binary complex, modulating the shape and charge of the ridge (Figure 3B). This conformational change enlarges the cavity, most likely to facilitate acceptance of an incoming primer strand (Figure 3B, right panel), in agreement with the catalytic advantage provided by a pre-formed preternary complex.

Replicative AEP primases are able to bind a "primer" as small as a single NTP, allowing the subsequent synthesis of a dinucleotide at the initiation step of RNA primer synthesis (Figure 6A). As bacterial NHEJ polymerases are closely related to these canonical AEPs, we tested the minimal length of DNA (incoming end) that can be accepted as a "primer" by *Mt*-PolDom (both DNA binary and preternary complexes) to determine whether these enzymes have a similar capacity to accept and extend off very short "primers," which in the context of NHEJ are short DSB overhangs. As shown in Figure 6B, three different molecules were tested as potential primers to be extended with CTP (the nucleotide preloaded in a preternary complex formed at a 3′-protruding [CCG] DNA end). The scheme on the left shows the degree of complementarity of the three primers used (3G, 2G, or G). Primer extension (p + CMP) was observed with 3G and 2G primers, but not with GTP, and was mostly dependent on the presence of the DNA end (preternary complex). Figure 6C shows that the minimal primer (2G) could only be extended when the DNA termini was complementary (CCG), and not with a noncomplementary end (TTG). Moreover, a complementary NTP (a requisite to form and stabilize a preternary complex) was required to form an extended product (dG-dG-CMP) (Figure 6C). Similar results were also obtained with primers matching a different DNA end (data not shown). Together, these results establish that formation of a preternary polymerase complex significantly enhances the extension off short incoming termini (Figures 6B and 6C).

## DISCUSSION

Conventional replicative and lesion bypass DNA polymerases extend off dsDNA substrates, containing both primer and template strands, in a 5′ to 3′ direction. In contrast, polymerases

Molecular Cell
Structure of a NHEJ Polymerase Preternary Complex

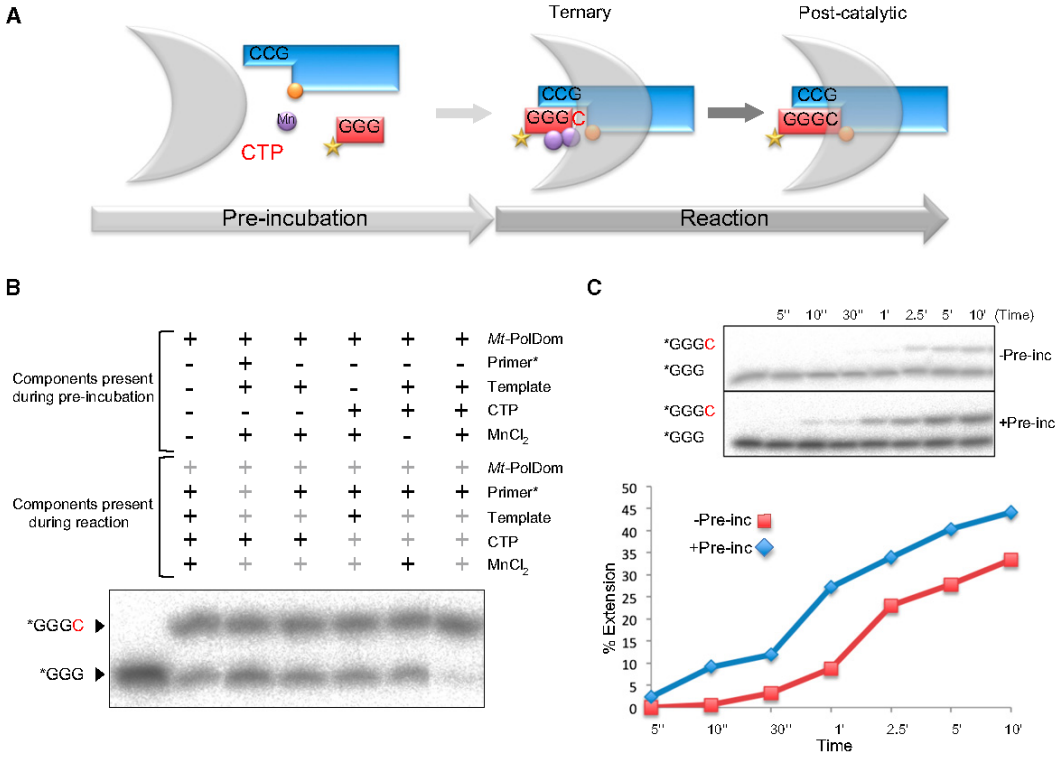



**Figure 5. Formation of a Preternary Complex Enhances Nucleotide Extension off an Incoming Primer Strand**

(A) Schematic representation of the experimental steps reported in parts (B) and (C) (below). Some of the components needed for the extension reaction (*Mt*-PolDom [gray crescent], a template molecule [blue] with a 5′ P [yellow dot] and a GCC-3′ protrusion, dGdGdG [red] as primer, CTP, and manganese ions) were allowed to interact during a preincubation step, as indicated in (B) for each case. Subsequently, the remaining components required for the reaction to occur were added, and the extension step can proceed.

(B) NHEJ polymerase extension reactions, following the scheme described in (A). Concentrations used were 600 nM *Mt*-PolDom, 1mM $MnCl_2$, 100 μM CTP, 5 nM labeled primer, cold CCG3′, and 25 nM template DNA. During preincubation for 30 min at 4°C, the components present are indicated by a + sign, while missing components are indicated by a – sign. At the start of the reaction period, added components are indicated by a black + sign, while components already present are indicated by a gray + sign. Reactions were performed for 30 min at 30°C, and the samples were run on a 6 M Urea-20% polyacrylamide gel and autoradiographed.

(C) Kinetics of NHEJ polymerase extension reactions following the scheme described in (A). –Pre-inc indicates preincubation of *Mt*-PolDom only; +Pre-inc indicates preincubation of the components needed to form the preternary complex (*Mt*-PolDom, nucleotide, template DNA, $Mn^{2+}$). Insertion reactions were stopped at the times indicated, and samples loaded on a 6 M Urea-20% polyacrylamide gel and autoradiographed.

involved in DSB repair must be capable of binding to and extending off noncanonical DNA polymerase substrates, including 3′-overhanging termini lacking a primer strand. Recent studies on the bacterial NHEJ polymerases have revealed some of the unusual activities associated with these repair enzymes that enable DNA extension under the most extreme conditions (Pitcher et al., 2007a). For example, a homodimeric arrangement of the mycobacterial NHEJ polymerases can facilitate the association of two incompatible 3′-protruding DNA ends, via microhomology-mediated synapsis, forming a stable end-joining intermediate (Brissett et al., 2007). This synaptic complex reflects an intermediate bridging stage of the NHEJ process, prior to end processing and ligation. In this way, the polymerase restores the continuity of the dsDNA helix, catalyzing a conventional 5′-3′extension reaction occurring on one DNA end, but templated in *trans* by a second (synapsed) DNA end.

Although recent studies have provided unique insights into polymerase-mediated orchestration of break synapsis, the order of substrate binding events and mechanism by which these NHEJ polymerases catalyze end extension is still poorly understood. To address this question, we have elucidated the crystal structure of a preternary intermediate of *Mt*-PolDom bound to DNA and NTP and shown that this complex is relevant for specific DSB repair processing events. This catalytically competent complex consists of a *Mt*-PolDom monomer, containing two metal ions and a templated nucleotide (UTP) in its active site, bound to a dsDNA end with a 3′ overhang but, significantly, lacking a primer strand (Figure 1). To our knowledge, this structure represents a unique example of a polymerase-DNA complex captured in a preternary intermediate state, relevant for NHEJ.

Is the preternary complex, described here, physiologically relevant for prokaryotic NHEJ polymerase extension reactions?

Molecular Cell
Structure of a NHEJ Polymerase Preternary Complex

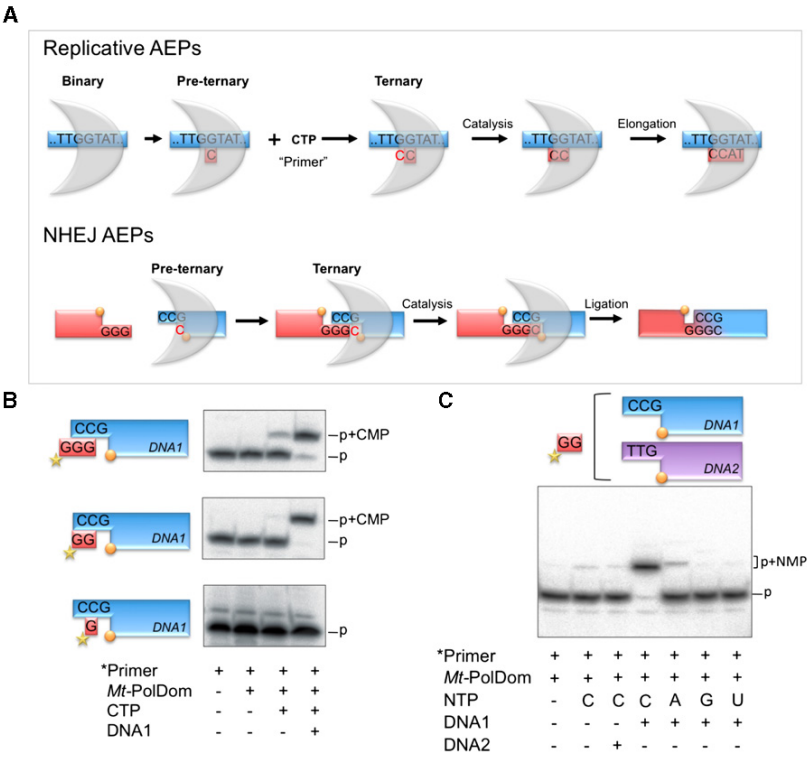



**Figure 6. Minimal Length of DNA that Can Be Accepted as a "Primer" by the Preternary PolDom Complex**

(A) A comparison of the reaction steps performed by replicative AEPs versus NHEJ AEPs. See the main text for details.

(B) A 3′-protruding (GCC-3′) DNA molecule (DNA1; 25 nM), with a 5′ P (yellow sphere) was used as the DNA end to form a preternary complex in the presence of 600 nM *Mt*-PolDom, 1 mM $MnCl_2$, and 100 μM CTP. Different DNA molecules (dGdGdG, dGdG, or dG; indicated in red), labeled at the 5′ end, were tested as primer strands (5 nM) to accept the CTP substrate preloaded in the preternary complex. After incubation for 24 hr at 4°C, primer extension products (p+CMP) were separated by 8M Urea-30% PAGE and detected by autoradiography.

(C) As in (B), but with two different DNA molecules (DNA1 and DNA2), both capable of forming a preternary complex with CTP, and different NTPs (100 μM) as indicated. The minimal primer (dGdG) could only be extended with CTP when a complementary template (DNA1) was present but not with an uncomplementary template (DNA2).

See also Figure S4.

Although the preternary complex lacks an incoming primer strand, which provides the attacking nucleophile (3′ OH), a comparison of the positioning of the nucleotide base, phosphate tail, active site ligands, and divalent metal ions to those in the active site of a ternary polymerase complex (Polλ) provides compelling evidence that the PolDom preternary complex is catalytically competent, even in the absence of the primer strand (Figure 2B). The possibility of preforming a preternary complex in solution by incubating the necessary components (*Mt*-PolDom, DNA end, complementary nucleotide, and activating metal ions) in the absence of a primer, allowed us to demonstrate its physiological relevance in accelerating NHEJ reactions, probably by providing a "ready to use" primer binding site. By testing the activity of the preternary PolDom complex with different single-stranded DNA primers, we concluded that the minimal primer utilizable by these enzymes is a dinucleotide, as *Mt*-PolDom was not proficient at polymerizing off a single nucleotide "primer" (Figure 6A). The fact that *Mt*-PolDom cannot efficiently use a single nucleotide as a primer indicates that although PolDom is evolutionarily related to replicative AEPs, its physiological activity as a primase has effectively been lost and, instead, these polymerases have evolved to have a more restricted capacity to bind short incoming DNA termini, enabling them to perform more specialized roles in NHEJ break repair processes.

From a mechanistic point of view, this study has identified a second conserved loop (loop 2), which plays a prominent role in activation/inactivation of the catalytic center. The conformation of loop 2 changes significantly upon the templated-binding of the correct incoming nucleotide, which induces the rotation of the $Arg^{220}$ side chain (~180°) away from the active site in the preternary complex (Figure 3A). Mutation of this invariant residue abolished the extension activity (Figure 3C) but, significantly, did not alter enzyme binding to other DNA substrates, such as gapped DNA. A comparison of the structures of the PolDom-DNA binary versus the preternary complex reveals the sequential movements that occur in the active site, induced by the binding of both a templating base and an incoming nucleotide (Figure 3A). The invariant active site residue $Phe^{64}$, which stacks against the base of the incoming nucleotide in the PolDom:GTP binary complex (Figure 2A), now stacks against the base of the templating nucleotide (A6) in both PolDom:DNA and preternary complexes, orienting this base and also maintaining (together with $Phe^{63}$) the major kink in the template strand (~105°). In replicative DNA polymerases, aromatic residues are employed as an essential part of a fidelity mechanism that scrutinizes pairing of the correct incoming base with the templating base, thus acting as a molecular gatekeeper to limit the incorporation of incorrect/mismatched bases during elongation (Johnson et al., 2003). We propose that an analogous fidelity mechanism involving the two invariant phenylalanine residues also occurs in the bacterial NHEJ polymerases, but in the absence of the primer strand, thus ensuring that the correctly templated incoming base is bound in the active site prior to the encounter with the incoming end/primer providing the attacking 3′ OH.

This phenylalanine-mediated ($Phe^{64}$) stacking interaction with the templating base in the preternary complex also promotes the movement of the incoming nucleotide (UTP) into the active site

Molecular Cell
Structure of a NHEJ Polymerase Preternary Complex

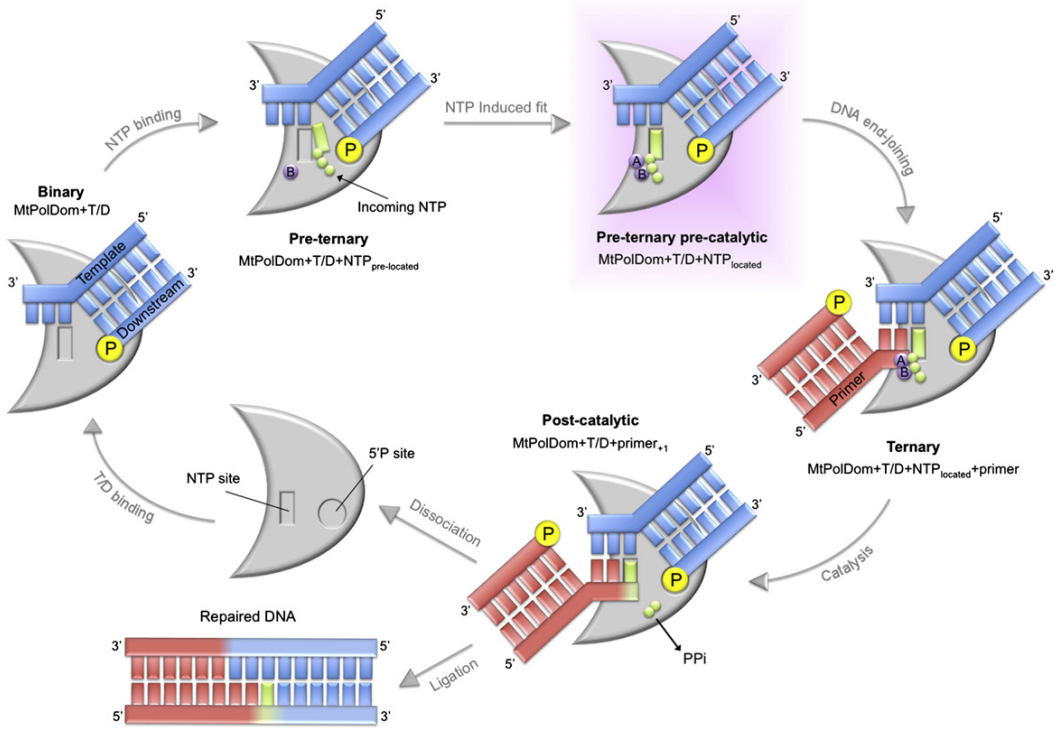



**Figure 7. Catalytic Cycle of a Prokaryotic NHEJ Polymerase**

Initially, a binary complex between the PolDom enzyme (gray crescent) and DNA (T/D; blue) is formed, mainly stabilized via interactions with the 5′ P. Binding of an incoming NTP (green) forms a preternary complex, still incompetent for catalysis as it lacks metal A. Upon template selection and relocation of the complementary NTP and the two metals, A and B, at the correct site (representing a primer-independent NTP induced-fit step) a preternary precatalytic complex is formed. This activated complex, described here, is ready for DNA end joining, allowing the 3′ OH of the incoming primer strand (red) to bind in the active site to form the ternary complex. Further steps of extension, PPi release, dissociation, and ligation (performed by the ligase domain of LigD), complete the DNA repair process.

and, together with the loss of specific contacts (e.g., $Arg^{246}$, $Lys^{175}$, and $Lys^{52}$) promotes the correct repositioning of the α phosphate group of the incoming nucleotide for catalysis (Figure 2A). This reorientated α phosphate moiety, together with $Asp^{139}$, forms a second metal binding site (A) not present in the binary structure, which is required for the two metal catalytic mechanism common to all DNA polymerases (Brautigam and Steitz, 1998). The binding of the second metal, in turn, promotes breakage of the salt bridge between $Arg^{220}$ and $Asp^{139}$, repositioning this aspartate into a catalytically favorable alignment with the other catalytic aspartates, the α phosphate group and the two bound metal ions, to form an activated preternary intermediate awaiting the arrival of the nucleophile (3′ OH of the primer strand) (Figure 3A). The primer-independent NTP induced-fit step is schematically shown in Figure 7. The catalytic incompetence of the R220A mutant highlights the importance of the interaction of $Arg^{220}$ with $Asp^{139}$. We propose that the maintenance of this amino acid pairing provides a significant barrier to catalysis until the enzyme becomes optimally bound to DNA, metals, and the correct incoming templated nucleotide. Once these are bound within the active site, a sequence of structural rearrangements promotes the binding of a second metal ion (A). The affinity of $Asp^{139}$ for this second metal promotes the loss of interaction with $Arg^{220}$, leading to expulsion of loop 2 from the active site, which results in full activation of the catalytic center. The movement of loop 2 away from the active site, most likely, promotes this activation step in two ways. The first consequence is that breaking the salt bridge is irreversible, leading to the release of the acidic side chain of $Asp^{139}$, which is involved in the binding of the second metal (A) within the active site, ensuring that it is optimally poised for catalysis. The second notable consequence, induced by the reorientation of loop 2, is a significant change in the ridge (Figure 3B) that surrounds the active site, which most likely allows the 3′ OH of the incoming primer strand to bind in the active site to form the ternary complex. Further steps of catalysis, PPi release, and ligation would terminate the NHEJ process. An illustration of some of the different complexes formed during the NHEJ cycle is depicted in Figure 7.

## Evolutionary Implications

Since the discovery of NHEJ in prokaryotes, there has been much speculation as to why AEP primases were selected as the primary NHEJ polymerases in bacteria. Prokaryotic NHEJ polymerases evolved, from primordial AEPs with an innate ability to make short RNA primers, into a distinct clade of adaptable

Molecular Cell
Structure of a NHEJ Polymerase Preternary Complex

end-joining nucleotidyltransferases capable of performing multifunctional roles required to process nonhomologous DNA ends during DSB repair. By comparing the structures and sequences of PolDoms with those of canonical replicative primases (e.g., *Pfu* primase) (Pitcher et al., 2007b), it is clear that both share a common primase catalytic core but there are several distinctive functional adaptations. These differences include the existence of loop insertions (e.g., loops 1 and 2) and a positively charged phosphate-binding region in PolDom, which distinguish the NHEJ AEPs as a distinct primase family (Figures S3 and S4). We have shown that the formation of a nonprimed preternary complex with the NTP preloaded is beneficial, as it facilitates catalysis to occur as soon as the incoming primer is recruited to form a ternary complex. This mechanism is reminiscent of the initiation reaction carried out by replicative primases (Figure 6A), in which the order of events is very similar: first, a binary complex between enzyme and DNA is formed, followed by a preternary complex in which the 3′ nucleotide is initially stabilized, followed by recruitment of the 5′ nucleotide (the "primer"), to form a ternary complex. Thus, as a reflection of their common phylogenetic origins, both NHEJ AEPs and replicative AEPs can catalyze an unorthodox addition of a ribonucleotide in the 3′-5′ direction, followed by (in the case of replicative AEPs) conventional (5′-3′) elongation events. The innate ability of AEPs to accept short primers may have influenced evolutionary selection of these enzymes by prokaryotes to become the NHEJ polymerase. Indeed, many bacteria encode additional AEP orthologs whose physiological roles remain to be determined.

In eukaryotes, PolX family members were adopted as the NHEJ polymerase of choice and the AEPs appear to be used predominantly in DNA replication processes. Is preternary complex formation also relevant for eukaryotic NHEJ polymerases? It has been demonstrated that human Polμ can catalyze NHEJ extensions on very short and incompatible DNA ends (Nick McElhinny et al., 2005; Davis et al., 2008), a reaction that can take advantage of a limited terminal transferase activity (Andrade et al., 2009) and that can occur with both dNTPs and NTPs. It is likely that formation of a Polμ preternary complex, as that described here, triggered by the strong recognition of a 5′-recessive phosphate and a reinforced avidity for the incoming nucleotide (both properties also intrinsic to Polμ) would be beneficial to carry out noncomplementary NHEJ of minimally processed ends in eukaryotes, although this remains to be proven.

## Concluding Remarks

Although a multiprotein NHEJ complex, including Ku and LigD, appears to be the major determinant for the gross physical association of DSBs (macrosynapsis), it is likely that processing enzymes (including polymerases and nucleases) within these complexes also play significant roles in orchestrating the association of the extreme termini of breaks (microsynapsis), especially nonhomologous ends, thus coordinating the processing of these termini prior to ligation. Although the structures of several PolDom-DNA complexes have provided "snapshots" of intermediates in the catalytic cycle of the bacterial NHEJ polymerases, the challenge now is to obtain structures of ternary complexes in the act of extending off the 3′ OH terminus of an incoming primer strand.

## EXPERIMENTAL PROCEDURES

### Crystallographic Analysis of the Preternary PolDom Complex

Crystallization, diffraction data collection, and refinement of the *Mt*-PolDom:DNA:UTP complex was performed as described in the Supplemental Results.

### DNA Substrates

PAGE-purified oligonucleotides were 5′ end labeled with [$\gamma$-$^{32}$P]ATP by polynucleotide kinase. The oligonucleotides used to generate the DNA substrates were the following: for gapped substrates, P15 (5′-TCTGTGCAGGTTCTT-3′), T32 (5′-TGAAGTCCCTCTCGACGAAGAACCTGCACAGA-3′), and 5′ phosphate-containing D16 (5′-GTCGAGAGGGACTTCA-3′); for NHEJ template/downstream (T/D) substrates, oligonucleotide 5′-CCCTCCCTCCCGCC-3′ (GCC3′) was used as template, hybridized to a downstream oligonucleotide 5′-GGGAGGGAGGG-3′ (DP), which contains a phosphate at the 5′ end. Oligonucleotides 5′-GGG-3′ and 5′-GG-3′ were used as primers. For EMSA, template oligonucleotide 5′-CCCTCCCTCCCGTT-3′ (GTT3′) was hybridized to DP to form a T/D molecule.

### Construction and Purification of *Mt*-PolDom Mutant Proteins

Site-directed mutagenesis (QuickChange, Stratagene) was performed on the overexpression plasmid for *Mt*-PolDom, DNA constructs were sequenced and transformed into *E. coli* B834(DE3)pLysS. Wild-type and mutant *Mt*-PolDom variants were overexpressed and purified as described (Pitcher et al., 2005).

### Amino Acid Sequence Comparisons

A multiple alignment of different primases and primase domains from members of the AEP family was done with the program MULTALIN (http://prodes.toulouse.inra.fr/multalin/).

### EMSA and Polymerization Assays

Assays were carried out essentially as described (Pitcher et al., 2007b). EMSAs were employed to analyze the interaction of *Mt*-PolDom with NHEJ intermediates in a final volume of 12.5 μl, containing 50 mM Tris-HCl (pH 7.5), 0.1 mg/ml BSA, 1 mM DTT, 4% glycerol, 5 nM labeled DNA, and different concentrations of *Mt*-PolDom. After incubation for 10 min at 30°C samples were mixed with 3 μl 30% glycerol and resolved by native gel electrophoresis on a 4% polyacrylamide gel (80:1 [w/w] acrylamide/bisacrylamide). For standard (gap-filling) polymerization assays, the incubation mixture (20 μl) contained 50 mM Tris-HCl (pH 7.5), 1 mM $MnCl_2$, 1 mM DTT, 4% glycerol, 0.1 mg/ml BSA, 5 nM gapped DNA, the indicated concentration of NTPs, and either wild-type *Mt*-PolDom or mutant R220A. After 30 min of incubation at 30°C, reactions were stopped by addition of loading buffer (10 mM EDTA, 95% [v/v] formamide, 0.03% [w/v] bromophenol blue, 0.3% [w/v] cyanol blue) and subjected to electrophoresis in 8 M urea-containing 20% polyacrylamide sequencing gels. After electrophoresis, the unextended and extended DNA primers were detected by autoradiography. NHEJ polymerization assays were carried out essentially as described above, but with independent DNA template molecules (unlabelled) and short homopolymeric oligonucleotides as a labeled primer, and specific reaction times, as indicated in the corresponding figure legends.

### Steady-State Fluorescence Emission Assay

Emission fluorescence data for 2-aminopurine (2AP) DNAs (ATD-Bio) and *Mt*-PolDom complexes were collected with a Varian Cary Eclipse fluorescence spectrophotometer. Samples were excited at 360 nm, and fluorescence emission data scans were collected between 340–470nm. Band pass slits were 10 and 20 nm for excitation and emission channels, respectively. Solutions containing 200 nM 2AP-labeled DNA, 200 nM *Mt*-PolDom, 50 mM Tris (pH 7.5), 150 mM NaCl, 10 μM $MnCl_2$, and 4% glycerol were titrated with stock solutions of UTP. Intrinsic protein fluorescence was subtracted. Fluorescence values were plotted and relative quenching was fitted to the equation, $F = ([UTP] / K_d [UTP]) + Ns [UTP]$. F is the relative quenching specific to the complex, and *Ns* is the nonspecific binding constant.

Molecular Cell
Structure of a NHEJ Polymerase Preternary Complex

## ACCESSION NUMBERS

The atomic coordinates and structure factors were deposited in the Protein Data Bank (ID: 3PKY; Table 1).

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Results and four figures and can be found with this article online at doi:10.1016/j.molcel.2010.12.026.

## ACKNOWLEDGMENTS

The A.J.D. laboratory is supported by grants from the Biotechnology and Biological Sciences Research Council and Medical Research Council. L.B. laboratory supported by grants from Ministerio de Ciencia y Tecnología (BFU2006-14390/BMC and CSD2007-00015) and Comunidad de Madrid (P2006/BIO306) and by an institutional grant to Centro de Biología Molecular "Severo Ochoa" from Fundación Ramón Areces. M.J.M. is a recipient of a contract from the Comunidad de Madrid. We acknowledge the European Synchrotron Radiation Facility for synchrotron radiation facilities and thank the beamline staff at BM16-1.

Received: December 21, 2009
Revised: July 9, 2010
Accepted: December 10, 2010
Published: January 20, 2011

## REFERENCES

Andrade, P., Martin, M.J., Juarez, R., Lopez de Saro, F., and Blanco, L. (2009). Limited terminal transferase in human DNA polymerase mu defines the required balance between accuracy and efficiency in NHEJ. Proc. Natl. Acad. Sci. USA *106*, 16203–16208.

Aravind, L., and Koonin, E.V. (2001). Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. Genome Res. *11*, 1365–1374.

Bowater, R., and Doherty, A.J. (2006). Making ends meet: Repairing breaks in bacterial DNA by non-homologous end-joining. PLoS Genet. *2*, e8.

Brautigam, C.A., and Steitz, T.A. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. Curr. Opin. Struct. Biol. *8*, 54–63.

Brissett, N.C., Pitcher, R.S., Juarez, R., Picher, A.J., Green, A.J., Dafforn, T.R., Fox, G.C., Blanco, L., and Doherty, A.J. (2007). Structure of a NHEJ polymerase-mediated DNA synaptic complex. Science *318*, 456–459.

Crooks, G.E., Ho, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: A sequence logo generator. Genome Res. *14*, 1188–1190.

Daley, J.M., Palmbos, P.L., Wu, D., and Wilson, T.E. (2005). Non homologous end-joining in yeast. Annu. Rev. Genet. *39*, 431–451.

Davis, B.J., Havener, J.M., and Ramsden, D.A. (2008). End-bridging is required for Pol mu to efficiently promote repair of non complementary ends by non homologous end-joining. Nucleic Acids Res. *36*, 3085–3094.

Della, M., Palmbos, P.L., Tseng, H.M., Tonkin, L.M., Daley, J.M., Topper, L.M., Pitcher, R.S., Tomkinson, A.E., Wilson, T.E., and Doherty, A.J. (2004). Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. Science *306*, 683–685.

Frick, D.N., and Richardson, C.C. (2001). DNA primases. Annu. Rev. Biochem. *70*, 39–80.

Garcia-Diaz, M., Bebenek, K., Krahn, J.M., Pedersen, L.C., and Kunkel, T.A. (2007). Role of the catalytic metal during polymerization by DNA polymerase lambda. DNA Repair (Amst.) *6*, 1333–1340.

Gong, C.L., Bongiorno, P., Martins, A., Stephanou, N.C., Zhu, H., Shuman, S., and Glickman, M.S. (2005). Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. Nat. Struct. Mol. Biol. *12*, 304–312.

Helleday, T., Lo, J., van Gent, D.C., and Engelward, B.P. (2007). DNA double-strand break repair: from mechanistic understanding to cancer treatment. DNA Repair (Amst.) *6*, 923–935.

Iyer, L.M., Koonin, E.V., Leipe, D.D., and Aravind, L. (2005). Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. Nucleic Acids Res. *33*, 3875–3896.

Johnson, S.J., Taylor, J.S., and Beese, L.S. (2003). Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. Proc. Natl. Acad. Sci. USA *100*, 3895–3900.

Kumar, S., Bakhtina, M., and Tsai, M.D. (2008). Altered order of substrate binding by DNA polymerase X from African Swine Fever virus. Biochemistry *47*, 7875–7887.

Mahaney, B.L., Meek, K., and Lees-Miller, S.P. (2009). Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. Biochem. J. *417*, 639–650.

Moeller, R., Stackebrandt, E., Reitz, G., Berger, T., Rettberg, P., Doherty, A.J., Horneck, G., and Nicholson, W.L. (2007). Role of DNA repair by non-homologous end joining (NHEJ) in Bacillus subtilis spore resistance to extreme dryness, mono- and polychromatic UV and ionizing radiation. J. Bacteriol. *189*, 3306–3311.

Nick McElhinny, S.A., Havener, J.M., Garcia-Diaz, M., Juarez, R., Bebenek, K., Kee, B.L., Blanco, L., Kunkel, T.A., and Ramsden, D.A. (2005). A gradient of template dependence defines distinct biological roles for family X polymerases in non homologous end-joining. Mol. Cell *19*, 357–366.

Pitcher, R.S., Tonkin, L.M., Green, A.J., and Doherty, A.J. (2005). Domain structure of a NHEJ DNA repair ligase from Mycobacterium tuberculosis. J. Mol. Biol. *351*, 531–544.

Pitcher, R.S., Brissett, N.C., and Doherty, A.J. (2007a). Non homologous end-joining in bacteria: A microbial perspective. Annu. Rev. Microbiol. *61*, 259–282.

Pitcher, R.S., Brissett, N.C., Picher, A.J., Andrade, P., Juarez, R., Thompson, D., Fox, G.C., Blanco, L., and Doherty, A.J. (2007b). Structure and function of a mycobacterial NHEJ DNA repair polymerase. J. Mol. Biol. *366*, 391–405.

Pitcher, R.S., Green, A.J., Brzostek, A., Korycka-Machala, M., Dziadek, J., and Doherty, A.J. (2007c). NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation. DNA Repair (Amst.) *6*, 1271–1276.

Weller, G.R., and Doherty, A.J. (2001). A family of DNA repair ligases in bacteria? FEBS Lett. *505*, 340–342.

Weller, G.R., Kysela, B., Roy, R., Tonkin, L.M., Scanlan, E., Della, M., Devine, S.K., Day, J.P., Wilkinson, A., di Fagagna, F.D., et al. (2002). Identification of a DNA non homologous end-joining complex in bacteria. Science *297*, 1686–1689.

Zhu, H., and Shuman, S. (2005). A primer-dependent polymerase function of pseudomonas aeruginosa ATP-dependent DNA ligase (LigD). J. Biol. Chem. *280*, 418–427.

Zhu, H., Nandakumar, J., Aniukwu, J., Wang, L.K., Glickman, M.S., Lima, C.D., and Shuman, S. (2006). Atomic structure and nonhomologous end-joining function of the polymerase component of bacterial DNA ligase D. Proc. Natl. Acad. Sci. USA *103*, 1711–1716.